EPFL

Thèse n° 10 732

Querying the Digital Archive of Science: Distant Reading, Semantic Modelling and Representation of Knowledge

Présentée le 1er mai 2024

Collège des humanités
Laboratoire d'Histoire des Sciences et des Techniques
Programme doctoral en humanités digitales

pour l'obtention du grade de Docteur ès Sciences

par

Alina VOLYNSKAYA

Acceptée sur proposition du jury

Dr F. Graezer Bideau, présidente du jury
Prof. J. Baudry, MER J.-F. Bert, directeurs de thèse
Prof. M. Krajewski, rapporteur
Prof. J.-F. Gauvin, rapporteur
Prof. S. Kenderdine, rapporteuse

École
polytechnique
fédérale
de Lausanne

2024

Résumé

L'archive des sciences est un lieu où les pratiques scientifiques sont sédimenteés sous forme de brouillons, de protocoles d'hypothèses rejetées et d'expériences ratées, d'instruments obsolètes, de visualisations dépassées et autres résidus. Aujourd'hui, alors que les sciences deviennent de plus en plus numériques, il en va de même pour leur archive, faisant émerger de nouvelles pratiques de recherche et ouvrant de nouvelles possibilités de connaissance pour l'historien (des *big data* à la longue durée). Ces collections se distinguent nettement des lieux de mémoire traditionnels. Ce qu'elles stockent n'est pas un ensemble d'objets tangibles et authentiques, mais des données à traiter et à interpréter par des algorithmes et des logiciels. La manière dont les données d'archive sont situées, décrites et présentées à l'utilisateur est encadrée et médiée par les technologies et infrastructures numériques. Comment ces nouvelles infrastructures numériques fonctionnent-elles, et comment façonnent-elles notre relation avec le passé scientifique ? Que pouvons-nous apprendre sur les sciences du passé à partir de leurs résidus, lorsqu'ils deviennent numériques et se transforment en données ? Et comment ces collections peuvent-elles être rendues utiles aux historiens et au grand public ?

Je pars de l'idée que l'archive numérique ne se contente pas de stocker des vestiges du passé mais devient un agent actif de leur interprétation. Si tel est le cas, il nous faut explorer les limites, les conditions et les potentialités des interprétations qu'elle propose et qu'elle rend possibles. Cette thèse explore la manière dont nous comprenons et interprétons le passé des sciences à travers leur archive numérique, en se penchant sur ses modes de représentation spécifiques, ses méthodes de traitement du passé et ses mécanismes de transmission.

Basée sur un large corpus de collections scientifiques et mélangeant des approches quantitatives et qualitatives, cette étude élabore les éléments d'une ontologie humaniste (plutôt qu'orientée vers l'ingénierie) pour l'archive scientifique, traduisant des concepts et perspectives de l'histoire des sciences dans un langage informatique. En expérimentant avec les méthodes offertes par le numérique (*distant reading*, modélisation sémantique) et les interprétations qu'elles suggèrent, cette thèse réimagine l'archive numérique comme une manière de *fabriquer le passé* (des sciences).

Mots-clés : archives numériques, archives scientifiques, résidus des sciences, lecture distante, modélisation sémantique, représentation des connaissances, histoire numérique, histoire des sciences, nouveaux médias

Abstract

The archive of science is a place where scientific practices are sedimented in the form of drafts, protocols of rejected hypotheses and failed experiments, obsolete instruments, outdated visualizations and other residues. Today, just as science goes more and more digital, so does its archive, giving rise to new research practices and opening new frontiers of knowledge for the historian (from *big data* to the *longue durée*). These collections clearly differ from the traditional *lieux de mémoire*. What they store are not tangible and authentic objects, but data to be processed and interpreted by computer algorithms and software. The way archival data is situated, described and presented to the user is prefigured and mediated by digital technologies and infrastructures. How do these new digital infrastructures operate and shape our encounter with the scientific past? What can we learn about the science of the past from its residues as they go digital and turn into data? And how could these collections be made meaningful for the queries of both historians and the wider public?

I argue that the digital archive does more than store some remnants of the past; it becomes an active agent in their interpretation. For this reason, we need to explore the limits, conditions, and affordances of the interpretations it offers and makes possible. This dissertation probes into how we understand and interpret the past of science through its digital archive, focusing on its specific modes of representation, the methods of treating the past it offers, and its transmission mechanisms.

Based on a large corpus of scientific collections and mixing quantitative and qualitative approaches, the study assembles the elements of a humanist (instead of engineering-oriented) ontology for the scientific archive, transferring concepts and perspectives from the history of science into computational language. Experimenting with the methods offered by the digital (distant reading, semantic modelling) and the interpretations they enable, this dissertation reimagines the digital archive as a way of *making the past* (of science).

Keywords : digital archive, scientific archive, residues of science, distant reading, semantic modelling, knowledge representation, digital history, history of science, new media

Table of contents

Acknowledgments

This dissertation has been made possible through the support and encouragement of many. My deepest gratitude is reserved for my two supervisors, Jérôme Baudry and Jean-François Bert, for guiding this project carefully and thoughtfully from the initial proposal all the way through to its defense. I am deeply grateful to Jérôme for his ever-present support and advice; for our many thought-provoking discussions; for giving me just the right amount of freedom and also for challenging me with (lots of) puzzling and seminal questions, steering me toward new and unexplored research paths. I am also very thankful to Jean-François for broadening my horizons and sharing his intimate knowledge about French archival and critical tradition. My research greatly benefited from his insightful comments and practical advice.

I extend my gratitude to the members of the jury – Jean-François Gauvin, Florence Graezer Bideau, Sarah Kenderdine, Markus Krajewski – for their interest in my work and their time and consideration in reviewing it.

I owe a special thanks to Sarah Kenderdine, whose *Cultural Data Sculpting* class first made me think about digital archives and led me to CERN, where the initial concept for this dissertation was born. I also thank Jean-Yves Le Meur for his warm welcome at CERN and all the anecdotes he shared about its history and its (digital) archive. My sincere thanks go to Franco Moretti for showing me that it is possible to fuse humanities reflection with digital methods while still maintaining a humanities perspective.

This work came to fruition through the generous support of the Swiss National Science Foundation Doc.CH grant (P0ELP1_192402), which provides doctoral students with the opportunity to pursue their own projects. I am grateful for this opportunity.

I am thankful to all my colleagues in administration, especially Sylvie Thomet and Céline Cordey, for their help in numerous matters. To my colleagues in the laboratory: Ion Mihailescu, Yohann Guffroy, Nicolas Chachereau, Marianna Fenzi, Simon Dumas Primbault, Semion Sidorenko, Joël Praz, Arthur Émile, Alexander Dmitriev, Olya Nicolaeva. It has been a pleasure to share both friendly chats over lunch and a critical stance towards more or less everything in our reading seminars. Special thanks go to Ion for introducing me to the instruments in experimental psychology and showing how intriguing they can be. One of the chapters also benefited from his thought-provoking (as always) comments.

To my family and friends scattered across countries and continents and still always being there. Without their support, I would hardly have made it to the end.

Last but not least, this dissertation would not have happened without MM – my *sputnik*, coach, source of inspiration and discipline, – and without Oscar, a little grand theoretician who lent value to it all.

to BN, BT and DB

Introduction

> Large memory is in books and small memory is all about the little things
>
> Christian Boltanski

The creation of the new inevitably results in the multiplication of the obsolete[1]. In the ever-evolving landscape of scientific inquiry, every innovation supplants its antecedents, leaving a trail of discarded and antiquated ideas in its path. This cycle of "creative destruction"[2] forms the bedrock of the dynamics of change in science. In order for a new scientific paradigm to take over, as Thomas Kuhn (1962) described, old ideas, theories and concepts must be discarded and dismissed, converted to non-knowledge or "intellectual rubbish" – to take up Bertrand Russel's title (1972).

Yet it is not only ideas that get discarded. It is also the entire material environment that succumbs to "the slaughter-bench of history" (Hegel): obsolete scientific instruments and apparatuses, experimental protocols, laboratory notebooks and log-books, drafts and photographs, graphs and notes, recordings of meetings and safety regulations. These material residues of science represent both by-products and scaffolds of the production of scientific facts. They are integral to the journey of scientific investigation and the shaping of its results, but remain invisible in the finished product that science offers to the world – the polished scientific *publication*.

Publications, ideas and theories, are retained in the scientific library, while the material residues get sedimented in the archive of science. In the scientific library, one finds what science has come to know about the world; in the scientific archive one encounters the remnants of *how* it came to this knowledge. The scientific library embodies the "grand" (and official) memory of science, chronicling its achievements, discoveries, and breakthroughs. The archive of science, preserving obsolete "scientific stuff", scraps, bits and pieces, represents rather the backstage and material imprints of the scientific process, all that remains "behind the scenes".

---

[1] On the philosophy of innovation and obsolescence see Lübbe 2003.

[2] The concept of "creative destruction" was developed by Joseph Schumpeter and is primarily known in the economic theory.

This dissertation focuses on such archives of science oscillating between sites of knowledge, memory sanctuaries, and dumping grounds. These archives are custodians of what was once at the forefront of novelty and innovation, but, through the forces of "creative destruction", was deemed obsolete, out of service, and consequently edged out of the scientific process. Such collections hold the tangible objects of memory, the remnants, alluding to all that which is no longer present, the entire material ensembles that have since vanished into obscurity. Yet they also constitute *lieux de savoir* (Jacob 2007; Bert, Ratcliff 2015) preserving objects of knowledge, integral to the fabric of historical scientific inquiry and shedding light on how science was conducted in the past.

This study explores different ways of knowing *with, through* and *from* the archives of science. In so doing, it focuses on a distinct chapter in the history of the scientific archive: its transition to the digital realm. It examines and interrogates *digital* archives, both those that represent digitized scientific residues and those that accumulate born-digital materials.

These collections clearly differ from the traditional sites of memory filled with tangible objects. What the digital repositories store is *data* to be processed and interpreted by computer algorithms and software. The way the data is represented, interconnected, and presented to the user is shaped and mediated by a variety of systems and technologies. The digital archive thus appears as a multilayered and complex technological infrastructure, embodying particular presuppositions about knowledge and memory. What are these underlying assumptions, and how do they shape our interpretation and understanding of scientific history? What can we learn about the science of the past from its residues and leftovers as they go digital and turn into data? These questions probe the ways in which digital archiving not only preserves but also recontextualizes historical scientific sources, offering new perspectives and interpretations of the past.

Querying the digital archive of science refers, firstly, to the evolving practices of engaging with digital collections, focusing on the novel methods by which scientific residues are stored, arranged, described and accessed in the digital repositories. Secondly, it offers an inquiry into the digital archive of science itself as a specific epistemic site, providing a vantage point from which to interpret the history of science.

This dissertation thus stands at the juncture of several turns: archival, digital, and material.

It resonates with the archival turn[3], shifting focus "from archive-as-source to archive-as-subject" (Stoler 2002, 93). While acknowledging the insights of archival information

[3] For the review and theorizations of archival turn(s) see Stoler 2002; 2008; Ketelaar 2017; Poncet 2019.

studies, this research primarily aligns with the broader humanities critical reflection on the nature of archives. It views the archive as an epistemological site that delineates the boundaries of what can and cannot be remembered[4]. In this vein, the study does not consider the archive merely as a storehouse of antiquities or a reservoir of information. Instead, it is approached as an active locus shaping the contours of knowledge, understanding, and memory. This perspective is further underscored by referring to "the archive" in the singular, as opposed to the more conventional plural "archives" often used in archival studies[5]. While the research devotes a good deal of attention to specific repositories and projects, it is also predicated on the premise that there is a distinct historical/technological configuration of the digital archive it seeks to delineate.

The study also engages with the ongoing dialogue about the relationship between history, historians, and the archive[6]. In this respect, it follows Michel de Certeau's argument that "the transformation of archival activity is the point of departure and the condition of a new history" (Certeau 1992 [1975], 75). It acknowledges the historicity of the archive itself, as articulated both by historians and archivists[7], while also examining how the historian's relationship with and to the archive evolves under digital conditions. It explores the transformation of "le goût d'archive" (Farge 1997 [1989]) – the historian's affinity and approach to archival research – in the digital era[8], looking at how historical practice is reshaped in the age of digital technology and how it in turn reshapes the archive.

Further, the project is situated within the digital turn, framing digital technology as a medium for memory and knowledge transmission. It operates on the hypothesis that the affordances and constraints of digital infrastructures profoundly influence users' interactions with and interpretations of archival artifacts. Thereby, the project aligns with a number of disciplinary fields focused on the digital landscapes that have seamlessly integrated into our

---

4 The two "grand" philosophical theories of the archive, those of Michel Foucault (1969) and Jacques Derrida (1998), conceptualize the archive not as a place but as a first principle, ἀρχή. Nevertheless, both conceptions have strongly influenced archival studies (see, e.g., Schwartz and Cook 2002). For their critique (on the part of the historian) see Steedman 2002. An overview of archival theories in the humanities is offered by Blouin and Rosenberg 2007; Manoff 2014.

5 On the semantics of the term see Got 2015.

6 For historians' reflection on their relationship with archives, see classical works by Arlette Farge (1989) and Natalie Zemon Davis (1990), as well as Burton 2006; Artières 2015; Anheim 2019. On the relationship between historian and archivist and the divide between them, see Blouin and Rosenberg 2011. For a discussion of this relationship from the archivist's perspective, see Cook 2011.

7 For a full-fledged "grand" history of the archive from the Middle Ages, see Friedrich 2018. On the history of scientific archives and collections, see Hunter 1998; Yale 2015; Keller et al. 2018. For examples of archival histories written by archivists, see Posner 1940; Duchein 1992; Cook 1997.

8 On "le goût d'archive" in the digital era, see a collaborative project by Frédéric Clavert and Caroline Muller (2018), as well as Liu 2018 reflecting on a "sense of the past" in the digital age.

daily routines. Firstly, it is informed by the field of media archaeology[9], with its emphasis on the material contours of storage and transmission. Particularly important for this study is the theoretical reflection by Wolfgang Ernst (Ernst 1999; 2013; 2015), who puts the digital archive at the center of his media archaeology program[10]. Complementing Ernst's perspective is the work of Bruno Bachimont, exploring the nature of the digital in relation to memory and preservation (Bachimont 2020; 2021). In considering the archive as a knowledge infrastructure, the study also draws on a wide range of research on digital (data/knowledge/information) practices and infrastructures, addressing indexing and classification practices (Bowker and Star 2000), metadata policies (Edwards et al. 2011; Gartner 2016), and knowledge representation (Bachimont 2007). In relation to memory technologies, this dissertation is informed by studies within the field of Digital Memory Studies, most notably those that elaborate on the issues of memory mediation and re-mediation (Dijck 2007; Erll 2012), including through the archive (Blom et al. 2016; Robertson 2014), as well as exploring the relationship of software to memory (Chun 2013) and articulating the question of "digital connectivity" (Hoskins 2017).

Lastly, the dissertation aligns with the "turn to things" that has been increasingly influential in social and historical sciences since the 1980s (Preda 1999). Attention to things is refracted in a variety of ways in anthropology (Appadurai 1988), Science and Technology Studies (Latour 2005) and history of science (Baird 2004; Daston 2007), philosophy (Bogost 2012; Coeckelbergh 2020), cultural studies (Brown 2001). Across these diverse academic perspectives, there is, however, a shared emphasis on recognizing the agency of material objects and their active role in shaping social order. Bruno Latour, for example, calls not simply to return to and re-engage with things themselves, but to bring them into the political realm (Latour 2005), to identify in them powerful agents and "the missing social masses" (Latour 1994), to give them a voice.

Focusing on the archive of residues that is solely concerned with things, glorifying and guarding things, this dissertation acknowledges the necessity to respond to this call[11]. It recognizes the active and dynamic role of "non-humans" in scientific practice[12] and, accordingly, their ability to provide testimony about it[13]. But more than that, it examines and problematizes the agency and the "voice" of things within the archive, analyzing how objects

[9] For the perspective of archaeology of new media in general see also Huhtamo 2011; Parikka 2013. On the material outlines of storage see Kirschenbaum 2012.
[10] Ernst's perspective as well as its alternatives are discussed in detail in Chapter 2.1.
[11] See also Volynskaya 2023.
[12] To complement and revise Latour's approach, see also de Boer et al. 2021.
[13] On things and technologies as witnesses, see Schuppli 2020.

are situated, represented, and contextualized; examining their interrelations with other objects as well as with people, concepts, and institutions; and ultimately assessing the archive's ability to "make" object-oriented histories.

Scientific residues and archives of science

While the humanities are believed to experience a *fever* towards the archive (Derrida 1998), science, in contrast, is often characterized by an absence of archival desire. As Thomas Kuhn noted, "For reasons and in ways that remain obscure to me, the sciences destroy their past more thoroughly than do mathematics or the arts" (Kuhn 1980, 190). From the perspective of epistemology, these "reasons" and "ways" might be traced to distinctions between the sciences and the humanities, which go back to Neo-Kantian and Diltheau philosophy. Within the classical dichotomy between *Naturwissenschaften* and *Geisteswissenschaften*, the archive, as the custodian of history (rather than nature), is clearly situated within the purview of the humanities engaged in an ongoing dialogue with its past. In contrast, science is often seen as rushing forward, fixated on the future, readily forgetting the past and not concerning itself with preserving it.

This image of a forgetful and forward-looking science has already been questioned within the history, sociology and anthropology of science. In the first instance, this shift in perspective has been driven by ethnographic analyses of scientific commemorative practices, offering a nuanced understanding of how scientific communities interact with their past. Notably, numerous studies look into everyday archiving practices inside the laboratory (Lefebvre 2013; Lefebvre et al. 2015), the relations between scientific and societal commemorations (Pestre 1999; Boudia et al. 2010; Lefebvre, Jolivet 2020), great names and the construction of their biographies (Söderqvist 2007; Shortland, Yeo 2008), disciplinary variations of memorial practices and their political and cultural uses (Abir-am 1998; 1999; Elliott, Abir-am 2000; Haddad 1999; Bosstraeten 2011). Geoffrey Bowker's *Memory Practices in the Sciences* (2008) stands out in this context, as he describes various configurations of memory practices in science in relation to technologies. His work illustrates how not only representations but also the technical infrastructures of memory evolve from one era to another, tracing the interplay between technology, memory, and scientific practice.

Another focus of the discipline is the history of scientific collections as a means of knowledge production. In this line of research, archiving and collecting practices are seen as a particular way of knowing (Pickstone 2000), and the scientific archive as a site of knowledge is positioned alongside the laboratory and the observatory. So, the life sciences can be characterized as a discipline where collecting, ordering, and classification practices are not just peripheral activities, but rather form the very core of its methodology (Strasser 2012; 2019). As shown by Lorraine Daston, one can even distinguish a distinct category of "sciences of the archive" encompassing disciplines from astronomy and geology to botany and genetics. These fields focus precisely on the "practices of collection, collation, and preservation conceived as an intrinsically collective undertaking" (Daston 2012, 162). Daston's *Science in the Archives* (2017), probably the richest study considering archives as a preoccupation of science, brings together, through the concept of the archive, various scientific practices across disciplines and historical periods, ranging from ancient astronomy to today's data mining, and traces the "affinities and continuities" between them.

Constituting what might be called the archival turn in the history of science, these studies focus on those collections that science itself recognises as the carriers of knowledge – repositories "of what a discipline considers worth knowing and preserving" (Daston 2017, 2). Yet, there exists another, less recognized, archive of science, one that gathers what a scientific discipline deems *unworthy* to preserve. Made up of outdated odds and ends of scientific routine, it might include a variety of artifacts from drafts to instruments, notebooks to photographs, experiment protocols to cafeteria menus. Fragmentary, incomplete and not anticipated for future use, these archives are not the result of purposeful collection for "an imagined community of disciplinary descendants" (Daston 2012, 164). Rather, they emerge from processes of exclusion and obsolescence, a filtration of what is no longer relevant, and are often preserved by mere chance.

As such, they offer an alternative perspective on the history of knowledge. Examining such archives, in contrast to analyzing publication libraries and data collections, resonates with Carlo Ginzburg's "evidential paradigm" – "a method of interpretation based on discarded information, marginal data, considered in some way significant" (Ginzburg 2013 [1989], 101). This approach shifts focus to the neglected or seemingly trivial data, in our case, to those artifacts that science itself does not regard as sources or carriers of knowledge worthy of preservation.

I propose to call such an archive *the archive of scientific residues*. Much like *wonders*, *souvenirs*, *curiosities*, or *antiquities*, the concept of *residues* does not speak to the intrinsic nature of things. Instead, it represents a relative category referring to certain practices of treating and valorizing objects. *Residues* describe what once played a role in the scientific process but eventually was recognized as out of service, disused, obsolete and excluded from the established framework of scientific inquiry.

At some point, however, the residues might be reappraised and come to be recognized as valuable traces of past scientific endeavors. Anthropologist Michael Thompson (2017 [1979]) termed such a dynamic of attributing value the "rubbish theory". According to Thompson, cultural artifacts can generally be categorized into two primary types: transient objects, which diminish in value over time, and durable objects, whose value remains stable or increases. Beyond these, however, Thomson identifies a third, often neglected category: rubbish. These are objects deemed valueless and non-functional. According to Thomson, this "rubbish" category escapes typical control mechanisms, enabling objects to transition from being transient to durable, essentially moving from being considered worthless to valuable (Ibid, 9). This shift occurs when an object's use-value is deemed non-existent, opening the door for it to gain new forms of value, such as historical or aesthetic worth.

It is precisely their loss of functionality[14] that turns these scientific bits and pieces into vessels of memory, sources of nostalgia, and subjects of aesthetic enjoyment. For example, an inoperative scientific instrument, adorned with the patina of age, might acquire an "age-value" and be re-evaluated as a piece of scientific heritage. In a similar vein, CERN old collider slides, heavily marred by mold, gain aesthetic worth and are showcased as art objects in exhibitions across Geneva and New York (Fig. 1)[15].

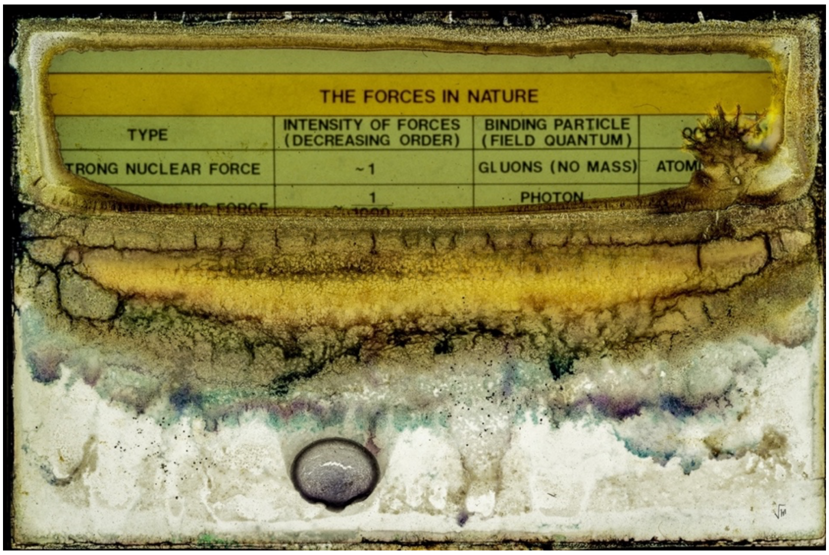


Figure 1. "The forces in nature". The VolMeur Project.

[14] As anthropologist Octave Debary observes, "These objects [residues]… become objects of memory from their 'defunctionalization': they are no longer preserved for their function, but precisely because they no longer have one (Debary, 2019, 66). My translation from French.
[15] The VolMeur Project, CERN Digital Memory artworks [https://volmeur.web.cern.ch/collection]

Apart from acquiring a newfound aesthetic and age-value, the archival artifacts can also attain a new semantic (heuristic) value[16]. Once in the archive, these very heterogeneous residues emerge as clues, traces by which a historian can retrace scientific practices of the past and the public can look behind the curtains of the scientific process. Items such as experiment protocols, laboratory notebooks, and technical drawings, once direct representations of nature, now serve as windows into the historical practices of representation themselves. They no longer point to the natural phenomena they were intended to document, but offer a lens into the methodologies and worldviews of the times in which they were created. Discarded by science as no longer functional or bearing knowledge, the residues of science might undergo a process of re-valorization and re-emerge as objects of historical knowledge.

However, whether such re-valorization occurs, and its specific manifestations hinge on the social, historical, and technological regime governing the archive. In this regard, the digital archive introduces a distinct modality, shaping the ways in which scientific residues are encountered and interpreted by historians, the broader public, and the scientific communities themselves.

Scientific residues go digital

Today, just as science goes more and more digital, so does its archive. The transformation of physical archives into digital formats has emerged as a cornerstone in the memorial strategies of scientific institutions (Lourenço, Wilson 2012). The scope of digitization initiatives, ranging from local projects to expansive national and international programs, is remarkably extensive (Ogilvie 2016). Some scientific institutions have not only digitized their physical collections, but also begun to update their digital archives on a daily basis, thereby incorporating archival practices into "scientific everyday life" and "turning all present data into archival entries and vice versa" (Ernst 2013, 99). The drive to gather and safeguard born-digital materials adds another layer to the scientific archiving landscape.

Beyond the conventional archival strategies of scientific bodies, the establishment of digital archives and databases for projects in computational history (of science) has significantly expanded the archival horizon. This expansion ranges from exhaustive digital collections of famous scientists' personal papers (such as *the Newton Project*, *Darwin Online*,

[16] While aesthetic value and age-value have more to do with the "patrimonial turn" (Hartog 2003; Boudia et al. 2010), heuristic value is more related to the history of science's shift toward the study of material everyday life of science. For a more detailed discussion cf. chapter 1.3

and *The Collected Papers of Albert Einstein*) to highly specialized databases dedicated to particular research areas (like the *Database Machine Drawings*).

In the digital archive, the residues of science (along with many other residues) are made *visible* as they never were before. Once consigned to the hidden depths of cabinets and storerooms, they now come to the fore, surfacing into the public view and gaining accessibility that was hitherto unimaginable.

In these new digital repositories, our interaction is no longer limited to discrete, individual artifacts; instead, we engage with their vast aggregations. Digital archives introduce new economies of scale unveiling *quantitative* landscapes of residues woven into the tapestry of "big data". While the traditional tangible archive is built on the process of selection of documents that are worth preserving, digital archives are driven by the process of accumulation. Instead of carefully discriminating between documents, they aggregate a wide variety of different materials, formats and media. As a consequence, the pre-digital concept of the archive, which "begins with selection" (Derrida 1998 [1995], 23), is being replaced by the vision of the total archive, referring back to the utopian projects of universal knowledge repositories, be it Paul Otlet's Mundaneum or Tim Berner Lee's Semantic Web.

In the history of science, the digital turn is also recognised as a crucial step to make this old dream come true: to establish a (total) archive for each discipline if not the entire history of science. A number of scholars emphasize that for the history of science to take advantage of the computational turn, it is essential to establish a centralized data archive and search system (e.g. Laubichler, Maienschein, Renn 2013; 2019; Siebold, Valleriani 2022), in which data is converted into a standard format and is equipped with a common metadata system. Seeking to overcome the fragmentation of data, such a repository is meant to aggregate "different types of data, revealing networks of connections and enabling comparison across far more types of data than would otherwise be possible" (Laubichler, Maienschein, Renn 2013, 122-123).

While digital collections promise to open new frontiers of knowledge for the historian, they are also altering the very notion of the archive. This redefinition extends to the concepts of storage, preservation, and user's interactions with archival artifacts.

In the digital archives, material scientific objects are transmuted into bitstreams to be read and interpreted by pieces of hardware and software. The direct tactile intimacy of engaging with the past, the materiality and authenticity of historical traces, are supplanted by new forms of technology-mediated (historical) imagination. In the digital archive, artifacts that once held the fascination of antiquarians, their surfaces bearing the patina of age and dust (Steedman, 2002), undergo a transformation into digital representations. The material residues of science are encapsulated in digital images and 3D models that invite virtual manipulation and "digital engagement".

The transformation wrought by technology extends beyond the mere digitization of archival objects; it fundamentally reconfigures the architecture and the very *arché* of the archive. Archival records are no longer governed by the traditional archival principles like provenance or original order. Instead, they are woven into and mediated by complex systems of knowledge organization, representation, and information retrieval. What users see on their screens, the methods they use to locate and extract "information", and the manner in which archival artifacts are represented, arranged, and interconnected, are all intricately shaped by multiple layers of systems and technologies. This complex network encompasses a wide array of components, from repository and content management systems to metadata standards, ontologies, controlled vocabularies, and interactive user interfaces (O'Neill, Stapleton 2022).

Engaging with information systems and digital technologies, this study still firmly anchors itself in the traditional humanities perspective. It examines technical systems and infrastructures in terms of the interpretations and representations they make possible. It probes into how the residues of science are (and can be) *contextualized* and imbued with *meaning* within the technological order of the digital archive. It explores the regimes of interpretation and memory, ultimately contemplating what it means to *understand* within the context, or under the mediation, of digital archives.

Addressing these questions, this dissertation is anchored in the constructivist historiography[17], which asserts that the past is not given to us *per se*, as a fait accompli, but rather is continuously re-invented and re-constituted. This line of research explores how the past is mediated, represented, and shaped through various regimes and forms of historical representation. So, Hayden White famously pointed to the importance of the literary/stylistic side of historical narratives, exploring how history takes shape through rhetorical tropes[18]. Stephen Bann extended the rhetorical analysis beyond historical narratives, applying it to the analysis of what he called the "historical poetics" of museums and collections. By analyzing the structure and forms of display in various collections[19], Bann explored the codes through

[17] On constructivism in the history of science, cf. Golinski 2008.

[18] As White claims, "it is the types of figurative discourse that dictate the fundamental forms of the data to be studied. This means that the *shape* of the *relationships* which will appear to be inherent in the objects inhabiting the field will in reality have been imposed on the field by the investigator in the very *act of identifying and describing* the objects that he finds there" (White 1986, 95). Cf. also Ankersmit 1988. On literary technologies as a way of "inventing" and communicating science, see Shapin 1984.

[19] Bann's most well-known case study is an analysis of the Cluny Museum's historical exhibit (Bann 1986). For other cases and some methodological reflection, cf. Bann 1990, 1994.

which history had been mediated and the ways in which these codes had been changing. The representation of history in museum thus stood for Bann as itself a historical problem indicating the accepted practices and modalities of dealing with the past. Along the same lines, this study approaches the digital archive, looking at how the archive in its current historical and technological configuration represents and thereby "makes" the past.

The idea that the past is invariably mediated by the means through which we access it forms a crucial tenet of this approach. Although historians, scientists, and the general public may each have different perspectives, strategies, questions, and interests when dealing with the digital collections, what is common in their engagement with the past is the way it is mediated and represented through the archives. It is precisely this common ground that this study targets. As Stephen Bann notes,

> What are the implications of this cross-disciplinary view of historical representation? In my view, it is only by recognizing and identifying the codes through which history has been mediated…that we can hope to avert a final cleavage between the circumscribed world of the professional historian and the generalized regime of spectacle into which all forms of popular representation risk being assimilated (Bann, 1990, 3).

Consistent with this rationale, this thesis seeks to explore the digital archive as a form of representation of the past that often remains invisible to both the historian and the public, but prefigures their "encounter" with the past. As such, this perspective differs from the ethnography of the archive, which would entail studying the individual experiences of users, archivists, and historians. While such an approach is prevalent and actively used in museum, collection, and digital platform research (Dobreva et al. 2012; Sinn, Soares 2014; Rhee 2015), this dissertation opts for a different investigative framework. It queries the *affordances* of the digital archive as an "invisible infrastructure" or a medium representing the past.

Questioning how our understanding of the past is shaped through technologies, this research also aligns with the disciplinary focus of digital hermeneutics – a broad perspective referring back to the classical works on interpretation by Hans-Georg Gadamer and Paul Ricœur and further to the tradition of the 19th century. Originally conceptualized as a theory and methodology for text interpretation, hermeneutics has since broadened its purview. It first expanded to encompass culture at large and subsequently evolved to analyze scientific technologies (Ihde 1999) and further the realms of (digital) technology and media (Zundert 2016; Romele 2019; Fickers et al. 2022).

Digital hermeneutics, broadly defined, is predicated on the assumption that technologies have a hermeneutic power (Ihde 1990), that is, they mediate and transform our experience; through technology we "read" and interpret the world. If so, one needs to understand how mediations and the interpretations they offer are configured and how they operate. Don Ihde's project of "expanding hermeneutics" focused on the use of instruments in scientific praxis. However, as the digital increasingly permeates the work of historians, their perspective and the very craft of history are also becoming influenced by the technology-shaped representations. Digital hermeneutics, therefore, offers a critical toolkit for recognizing and reflecting on this influence.

This thesis endeavors to develop a hermeneutical approach for the digital archive of science, employing and experimenting with its two "methods": distant reading and semantic modelling. Distant reading was conceptualized by Franco Moretti (2013) as a method of approaching an archive of forgotten books, or as Moretti (2000) puts it, "the slaughterhouse of literature"[20]. In terms of specific methodologies, it involves a suite of text mining techniques ranging from frequency and keyword analysis to topic modelling and network analysis. Through these methods, distant reading allows for a macroscopic exploration of textual data, uncovering latent patterns, connections, and trends not immediately visible through traditional, close reading methods. Originally framed for (literary) texts, distant reading has become part of "a broad intellectual shift that has also been transforming the social sciences" (Underwood 2016, 530). Adapting the technique of distant reading to the realm of scientific collections, I will examine "the archival order" spanning across various collections and repositories. This analysis will shed light on the commonalities and differences in the ways scientific residues are represented, curated and contextualized (*Part II*).

Semantic modelling as a method involves constructing a model (an ontology) that organizes the relationships and meanings within the data, transforming raw archival content into a machine-readable and interconnected "web of knowledge" (Peckham, Maryanski, 1988; Hawkins 2022). As an "experimental" part of the research, I will construct an ontology for contextualizing the residues of science within the archive and animate it with a case-study (*Part III*). Both of these methods, albeit from different perspectives, aim to articulate, problematize and question the modes and forms of (knowledge) representation employed by the digital archive.

The adopted humanities stance is more than just a selected methodology among many—it embodies a foundational premise of the research. The study is driven by the idea that digital archives of science (like many other infrastructures) not only demand the insights of information managers, archivists, and knowledge engineers, but also necessitate an

---

[20] In a sense, the archive of scientific residues appears as a variation of the "slaughterhouse of science"

interpretative and contextual expertise offered by the humanities. As both historians and the general public are increasingly engaging with history through technological mediation, the question extends beyond efficiently storing and retrieving information. If the digital archive offers distinct modes and frameworks for representing the past, we need to understand how these historical representations are configured and how they function; what becomes visible through the digital archive of science, and what, on the contrary, remains hidden. What is at stake here is how we engage with the history of knowledge (and science), how we can process, understand, and make sense of it, and what we can learn about it through digital archival technologies.

This study neither aligns with techno-optimistic narratives that champions digital systems as pathways to universal models of knowledge, nor does it embrace techno-pessimism that views technology as a threat to both history and the archive. Without treating the digital as the answer to all questions, it proposes to critically investigate the epistemological possibilities of the digital archive, exploring the (new) methods it offers, and their capacities and limitations.

Articulating these questions, this dissertation contributes firstly to the epistemological and methodological reflection of the computational history of science. It probes the methods and boundaries of using the digital archive of science, as well as its capacity to bring forth new meanings, interpretations and new knowledge. The research therefore seeks to provide a critical understanding of the infrastructures that the historian of science deals with in the context of the digital turn.

Secondly, the study elaborates on the problematics of preservation of science, discussing how to store, contextualize, and represent scientific residues as objects of knowledge that can shed light on the science-in-the-making of the past. In this respect, the project contributes to the debates on the role of things as actors of the scientific process and its witnesses. It also enriches both theoretical discussions and practical applications in the field of digital preservation.

Thirdly, addressing the questions of scientific memory and its transmission through the digital archive, this dissertation not only contributes to the theoretical reflection within the realm of Digital Memory Studies but also offers some insights for enhancing the archival and memorial practices of scientific institutions.

Structure of the dissertation

This dissertation probes into the digital archive of science focusing on its peculiar modes of representation, the methods of treating the past it offers, and its transmission mechanisms. These three focal points shape the structure of the study aligning closely with its three main parts (II, III, and IV).

Part I sets the scene for the ensuing discussion. It introduces and shapes the concept of the digital archive of scientific residues, outlining the main lines of discussion and the key tensions inherent in the concept. It commences by contrasting two ideal-typical models: the "traditional archive" and the "knowledge repository", each characterized by its distinct techniques, principles, and terminologies. I then introduce the digital archive into this dichotomy, demonstrating how it maintains a tension with both models. Charting the contrast between traditional and digital archive configurations, I frame the digital archive within a technological horizon, describing it as a distinct *dispositif*[21] grounded in the material realities of record production, record-keeping, and storage. Then, through a close reading of one of the key standards of digital preservation – the OAIS model – I demonstrate how the very conception of archiving and preservation, alters from the traditional archive to the digital one. As I show, the OAIS framework introduces both pragmatic and hermeneutic reformulation of the archive, a concept that will be revisited and further explored in subsequent parts of the dissertation.

In addition to the technological aspect, this part also sets out the historical context for the archive of scientific residues. As I demonstrate, the archive of scientific residues "comes into being" and becomes possible when Science studies get interested in the everyday life of science and the role of the thing in it. Thus, the digital archive of science, as conceptualized in this dissertation, is situated within and emerges from the concurrent interplay of specific historical and technological horizons. The introductory part concludes with a (rather broad)

---

[21] Here and hereafter, in speaking of *dispositif*, I am referring to Michel Foucault's concept – a "thoroughly heterogeneous ensemble" of both discursive and non-discursive elements, that structures and exercises power relations (Foucault 1980, 194-195). More precisely, I follow an extended interpretation of the concept by Giorgio Agamben, who proposes to term a *dispositif* "literally anything that has in some way the capacity to capture, orient, determine, intercept, model, control, or secure the gestures, behaviors, opinions, or discourses of living beings. Not only, therefore, prisons, madhouses, the panopticon, schools ... and so forth (whose connection with power is in a certain sense evident), but also the pen, writing, literature, philosophy, agriculture, cigarettes, navigation, computers, cellular telephones and—why not—language itself" (Agamben 2009, 14). By speaking of the digital archive as a *dispositif*, I refer precisely to its capacity to structure, regulate, mediate, and shape our relationship with the objects it represents. I mean that the digital archive (of scientific residues) constitutes a particular technological, social, cultural, and political configuration that entails and makes possible certain forms of representation, certain forms of engagement with artifacts of the past, and certain models of knowledge.

definition of the digital archive of scientific residues and a description of the archive corpus to be analyzed in the following part.

Part II delves into the representation of scientific residues within the digital archive. Treating the digital archive as a medium in its own right, it explores how it shapes our encounter with the residues of science, and, through them, with the scientific past. How do digital archives situate, contextualize, and assign meaning to scientific objects? Drawing on a corpus of over a hundred collections, I provide a detailed account of the morphology and architecture of the digital archival order: its organization(s), visual regime(s), structures of links it establishes and classification languages it adopts. Through both distant reading of the corpus and close-reading of the particular collections, I show how this new order challenges the traditional forms of historical representation and imagination. Another conclusion concerns the existing practices of representation of scientific residues. The analysis uncovers that while existing digital collections make the residues visible, searchable, and retrievable, they do not necessarily render them readable or open to interpretation. It thus identifies a disparity between the accessibility of objects and accessibility of their interpretation and context.

In the second half of the thesis, I put forward some reflections on how to make sense of a scientific residue in/through the digital archive. Parts III and IV present two case studies, addressing two generically different types of archives and the methods of dealing with them.

Part III picks up the question of archival representation, asking how to make residues of science readable and accessible for the history of science. It responds to this question through a peculiar experiment: *modelling* the archive. In this part, I present an ontology for the archive of scientific residues aimed at preserving and re-using such artifacts as objects of knowledge. The ontology offers three models for describing a scientific artifact: "biography" (tracing production, circulation and usage of an artifact), "assemblage" (highlighting how an object enters into relations with other objects), and "mediation" (focusing on the connection between the object and the phenomenon under study). Each of these models is grounded in certain interpretative approaches to things and offers a distinct language for contextualizing them.

I then exemplify how the ontology works through a case study of one minor device from the history of experimental psychology – the reaction key. Investigating how the reaction key can be contextualized through different models, I assemble its "archive", intertwining archival fragments across collections and repositories and explore the connections, interpretations, and histories that the model can generate. This part thus pursues two aims: first, it probes into semantic modelling as a way of contextualizing or "giving voice" to the residues of science within the archival framework. Second, it reflects on the *modelling* as a historiographical method and as an epistemological operation, examining its potentials, limitations, and boundaries.

Part IV centers on the *born-digital* residues of science and explores how these are transmitted into the future. While the previous part focuses on modelling the archive by meticulously connecting scattered archival fragments through digital technologies, this one ventures into a completely different case study. It delves deeply into one of the first born-digital archives, the Stanford Artificial Intelligence Laboratory's archive, SAILDART. Preserved somewhat serendipitously and not initially intended as an archive, SAILDART has retained almost the entire contents of the laboratory's computers from 1972 to 1990. Using the example of this archive, encountered nearly half a century after its creation, this part questions what does such a "total" archive transmit over time, and how.

Tracing archive (media) archaeology and mediology, I outline a peculiar configuration of memory it offers. In it, the computer emerges as the carrier of memory, its mediator, and even the subject of memory work. I further articulate and explore this form of machine memory and outline the ways in which it reflects the history of the laboratory – one of the pioneers of AI. In addition, I also pose the question of how to deal with this kind of memory: how to understand digital fragments of the past many of which are written in different languages for different systems that are no longer in use? The chapter delineates some methods for archival exegesis while also probing the epistemological possibilities of such an archive that resides in the liminal space between memory and oblivion.

The conclusion summarizes the study by bringing together the themes of archival representation, modelling, as well as transmission, and discussing both conceptual and practical implications of the analysis.

Part I. Digital archive (of science): boundaries and tensions

From the image of the Borgesian library to the concepts framed by Michel Foucault and Jacques Derrida, the archive has turned into a conceptual metaphor of culture that defines the conditions and limits of knowledge, what can or cannot be remembered and preserved. Yet *the archive* equally refers to actual institutions, as well as to the subject of archivists' care, and to the contact zone, where historians come in touch with the records of the past. Archivists tend to speak about an *archive*, while humanities scholars, in the course of their critical analysis, rather refer to *the Archive* (Manoff 2004). Digital Humanities introduce the "digital archive", pointing to specific projects (Gerber, 2023), whereas the field of media archaeology views it more as a specific technical configuration (Ernst 2012). Not to mention the numerous archival turns (Bastian 2016; Poncet 2019) across disciplines, each redefining the concept in its own way.

Without aiming for a precise and formal definition, this part will delineate some coordinates to frame the subsequent discussion on the digital archive of scientific residues. Situating the digital archive, I will start with the "traditional" one. Chapter 1.1 describes the (ideal-typical) model of the traditional archive and contrast it with the informational (documentary) model. I will then bring the digital archive into this contrast, outlining its points of tension and alignment with both the archival and the informational paradigms. Chapter 1.2 looks at how the digital archive is given a form in/through the principal standard of digital preservation – the OAIS reference model. Subsequently, in Chapter 1.3, I shift the focus to the interplay between the archive (of science) and the historian. I argue that it is within this relationship that the "archive of residues" comes into being. Finally, Chapter 1.4. summarizes some of the previous discussion and presents a *corpus* of archival collections that is analyzed in the following parts.

1.1. Traditional archive, knowledge repository and the digital

1.1.1. Traditional archive: a document as seen by archivists

> A single document out of a Group of Archives is no more to be taken as expressing in and by itself all it has to tell us than would a single bone separated from the skeleton of an extinct and unknown animal.
>
> Sir Hilary Jenkinson[22]

In 1898, three Dutch archivists, Samuel Muller, Johan Feith, and Robert Fruin, published the *Manual for the Arrangement and Description of Archives*, known in archivistics as the *Dutch Manual*. "A bible of archivists", the *Manual* sets out one hundred rules of archiving, many of which remain the core principles of archival science to this day[23].

The first rule, "the foundation upon which everything must rest," defines the archive (*archief*) as

> the whole of the written documents, drawings and printed matter, officially received or produced by an administrative body or one of its officials, in so far as these documents were intended to remain in the custody of that body or of that official (Muller, Feith, Fruin 1968 [1898], 13).

This classic definition outlines an exclusively paper-based bureaucratic archive with only government or corporate records being the subject of preservation and care. Excluded from archival custody are, inter alia, personal collections[24] and material objects other than paper documents.

The traditional archive is originally conceived as an "in-house" organization, inextricably linked to administrative bodies, mirroring and safeguarding their respective bureaucratic logics. It thus implies a particular perspective on the past, grounded in the record-keeping practices of institutions and administrative bodies. It represents a bureaucratic view of history, focusing on grand political narratives as seen through the lens of administrative bodies, in counterpoint to social or micro-history.

---

22 Jenkinson 1948, 4.

23 On the history of archives prior to the 19th century, see Duchein 1992; Yale 2015.

24 While the understanding of archives has significantly expanded, private collections are still not commonly referred to as archives. Instead, they bear distinct titles such as "manuscript collections", "papers", "personal records", or "private fonds". Until recently, it was believed that these collections should be treated differently from "archives" (Desnoyers 1991), and those in charge of them were referred to not as "archivists" but as "manuscript curators" (e.g. Evans et al. 1974).

Another, perhaps even more crucial, postulate introduced by the *Dutch manual* definition is the integrity and the organicity of the archive. Archival collection is seen as “an organic whole, a living organism” (ibid., 19), the structure of which reflects the operation of the administrative body in question. What the traditional archive preserves is the integrity of the body of documentation received from the donor. It is only in the context of this organic whole[25], formed in the course of record-keeping practices, that each specific document gains value.

The archive thus appears as a natural manifestation, an emanation of a corresponding organization. The archivist should intervene in its arrangement as little as possible. Instead, they are called upon to guard the integrity and organicity of its unique structure as it evolved within the administrative body[26]. Archival collections hence have nothing to do with *collecting*. As Hilary Jenkinson, the leading British archival theorist of the first half of the twentieth century, puts it,

> Archives are not collected: I wish the word “Collection” could be banished from the Archivist’s vocabulary …. They came together, and reached their final arrangement, by a natural process: are a growth; almost, you might say, as much an organism as a tree or an animal (Jenkinson 1948, 4).

The archive develops on its own, without the intervention of the archivist. This self-effacement of the archivist, echoes the famous self-effacement of the historian as expressed by Leopold von Ranke: letting things speak for themselves. The ethos of the archivist is to serve the ideals of *objectivity* and *impartiality*: the organization of the archive must not be subordinated to any particular type of user or any particular interest.

Since the *Dutch manual*, the mission of the archive has been to maintain the natural order of things, as opposed to the various forms of interference: re-ordering, re-configuring, re-modelling. The organic metaphor is still quite prominent in the archival discourse[27], which makes a distinction between “organic” collections formed “naturally” through bureaucratic practices, and “artificial” collections that bring together items from different origins (Duranti, Franks 2015, 33).

---

[25] Archives are still being defined in terms of wholeness. See, for example, the definition of “fonds” adopted by the General International Standard Archival Description (ISAD(G)): “The whole of the records, regardless of form or medium, o and/or accumulated and used by a particular person, family, or corporate body in the creator’s activities and functions” (International Council on Archives 2000, 10).

[26] As the Dutch Manual puts it, “In the rules that follow there is a careful avoidance of giving any scheme for archival arrangement and grouping” (Muller, Feith, Fruin 1968 [1898], 19)

[27] On the organic metaphor in archival science cf. Ilerbaig 2016.

At the level of practice, the ideal of archival integrity manifests itself in the three fundamental principles of archival science: *respect des fonds*, provenance[28], and original order (rules 8 and 16 in the *Dutch manual*). The principle of original order, "the most important of all" the rules, states that

> The system of arrangement must be based on the original organization of the archival collection, which in the main corresponds to the organization of the administrative body that produced it (Muller, Feith, Fruin 1968 [1898], 52).

The archive's role is not to impose a perfect order on the material or arrange it in a way that enhances its meanings. Instead, its purpose is to preserve the existing order, which is viewed as the ideal context for each individual record.

The *respect des fonds* rule states that "the various archival collections placed in a depository must be kept carefully separate" (Muller, Feith, Fruin 1968 [1898], 33). In other words, archival objects must be classified and kept according to their origins, i.e. by those bodies in which they originated. Various interpretations of the principle of provenance exist across different traditions and schools. They are implemented in diverse systems of archival organization, such as "fonds," "record groups," "archive groups" (Duchein 1983, Sweeney 2008).

These principles – standards of memory and recordkeeping, "definitive facts of archival science" (Duchein 1983, 66) – shape the very idea of the archival document, distinguishing it both from objects in "artificial" collections and from everyday things. An archival document is what belongs to the whole and gains meaning only insofar as it belongs to a whole. Only as part of the whole, within the general context of fonds, can it be assessed as authentic and unique. In this sense, the authenticity, uniqueness, and enduring value of archival objects are less attributes of the documents themselves, than a derivative of the configurations of the "traditional archive." As articulated by historian Étienne Anheim and archivist Olivier Poncet,

> the act of "archiving" [*mise en archive*] documents represents a deliberate operation that shifts the existing documentary logic to establish a new framework. This process assigns each document specific coordinates within series, collections, and locations, elements that subsequently become inseparable from the document itself (Anheim, Poncet 2004, 3).

[28] The principles of *respect des fonds* and provenance were adopted in practice in some countries even before the Dutch Manual. The principle of *respect des fonds*, associated with the French tradition, was introduced by the French Ministry of the Interior in 1841. The principle of provenance (*Provenienzprinzip*), more closely linked to the German tradition, was established by the Privy State Archive in Berlin in 1881. Subsequently, these principles were implemented in various ways across different national traditions. For a detailed exploration of the complexity in defining the principle of provenance and its distinction from *respect des fonds*, see Sweeney 2008.

It is upon crossing the "archival threshold" and entering into the system of archival conventions, standards, and principles, that a piece of paper turns into an archival document, "a permanent monument to its creator's actions" (Duranti 1996, 247).

The configuration of the traditional archive began to gradually destabilize starting from the 1960s. While the *Dutch manual* dealt with finite and manageable ancient collections, by the mid-century, the volume of documentation produced had already exceeded archival capacities. Archives could no longer preserve everything. In response to the "archival overload", the American archivist T. R. Schellenberg proposed that archivists should *appraise* records, i.e. assess the value of documents and make choices about what should be archived and what should be discarded, what should be forgotten and what should be remembered. According to Schellenberg, it is the archivist's appraisal that shapes the composition of an archive. The archive thus appears as a *selection* from the broader organic whole delivered by the administrative body.

The very idea of appraisal alters the role of the archivist from a custodian of a natural order to a more proactive agent, governing the processes of memory and oblivion. Sir Hilary Jenkinson, a prominent opponent of Schellenberg's approach[29], for instance, believed that the archivist should not intervene in the archive formation process. He advocated for the archivist to be an unbiased and objective custodian, guarding the organic integrity of the records:

> The Archivist's career is one of service. He exists in order to make other people's work possible. ... His Creed, the Sanctity of Evidence; his Task, the Conservation of every scrap of Evidence attaching to the Documents committed to his charge; his aim to provide, without prejudice or afterthought, for all who wish to know the Means of Knowledge.... The good Archivist is perhaps the most selfless devotee of Truth the modern world produces. (cited in Cook 1997, 23).

The introduction of appraisal as a mandatory procedure in archival practice began to destabilize the ideal-typical[30] construct of the traditional archive, which was predicated on preserving fonds "as they are". This movement towards the de-naturalization of the archive gained further momentum in the wake of postmodernism, which led to the reevaluation of many classical archival postulates, at least at the level of theoretical reflection (Cook 1994; 1997). Subsequently, and perhaps even more significantly, the emergence of digital technologies

[29] For further discussion of the debate, cf. Cook 1997, Duranti 1994.

[30] Using the concept of an "ideal type", I align with Max Weber's notion of it as an analytical construct that allows for the identification of regularities, similarities and differences among specific empirical cases. Ideal type, in other words, is an abstraction, functioning as a heuristic tool to understand (social) phenomena.

compelled archivists to rethink and reinvent many principles that, back in the early 1980s, seemed "never again to be fundamentally questioned" (Duchein 1983, 66).

Nonetheless, the core principles of provenance, original order, and the concept of a cohesive, organic collection remain at the heart of the archival discipline. These notions continue to serve as the benchmark for judging what constitutes an archive and how to distinguish between archive and non-archive. Notably, this line of argument is evident in debates surrounding the digital archive. Many archivists are hesitant to label certain digital collections as "archives," precisely because they do not conform to the traditional archival principles (a viewpoint articulated, in particular, by Kate Theimer (2012)).

1.1.2. Knowledge repository: a document as seen by Library and Information science

> All bibliological creation, no matter how original and how powerful, implies redistribution, combination and new amalgamations from what is previously given.
>
> Paul Otlet[31]

In 1895, three years prior to the publication of the *Dutch manual*, two Belgian lawyers, Paul Otlet and Henry La Fontaine, founded the International Office of Bibliography in Brussels and launched the Universal Bibliographic Repertory. This event is often considered to be the beginning of European documentation studies and, by extension, the dawn of information science (Rayward 1997, Day 2008).

"A set of techniques needed to manage the explosion of documents" of the late 19th century (Buckland 1997, 804), documentation science inherits the numerous information management practices that extend back for centuries (Blair 2011, Krajewski 2011). In turn, this field of study later informed Library and Information science (LIS) – an umbrella discipline encompassing information retrieval, knowledge representation, knowledge management and information literacy, infometrics and web science (Stock 2015, 9). The projects of Otlet, as well as those of his predecessors and successors, focus on how to organize knowledge[32] in a way that renders it viable, operable and accessible.

---

31 Otlet 1934, 422-423.

32 Hereinafter, referring to "knowledge" and "information", I attempt to follow the usage of the documentalists and/or information scientists. But since I am merging Documentation Studies and LIS into a single paradigm, it becomes challenging to distinctly delineate or define these concepts. The conceptualization and use of "knowledge" and "information" vary from one project to another, leading to a fluidity in how these terms are understood and applied within this combined framework. As for

Like archivists, documentalists and information managers deal with documents. But if the archival document is unique, authentic, and situated in its proper place within archival fonds, the document as seen by LIS, is informative, usable and integrated within a valid knowledge organization system, such as bibliographic repertory, indexing system, or database.

While the archival document gains its value within the configuration of "traditional archive," the document as a piece of knowledge gets its meaning within the "knowledge repository," as an infrastructure for knowledge organization, exchange and retrieval. As with the traditional archive, a *knowledge repository* is understood here as a particular (ideal-typical) mode of processing documents, which takes different forms in actual projects and implementations of libraries, document centers, databases and data banks.

Two processing modes are thus developing alongside each other: on the one hand, the "archival paradigm" placing the documents within the archival organic whole, and on the other hand, the "information paradigm", situating the documents in the knowledge repositories so that they can be retrieved and (re)used.

The difference between the two paradigms is often articulated as a difference in the nature of the material: archives are assumed to be concerned with the unique and unpublished, while documentation studies supposedly deal with the published sources of knowledge. Yet archives often keep printed documentation, which "gains" uniqueness and authenticity due to its singular place in the structure of the fonds. At the same time, for instance, a library may well hold handwritten manuscripts treating and situating them as pieces of knowledge, for instance, categorizing them according to thematic classifications.

The very boundaries of what is understood as a document in documentation studies are quite fluid (unlike in archives, which dealt exclusively with paper until the 1970s-80s). In Otlet's works, the concept of document is already extended beyond paper-based materials to a variety of media. This move is more explicitly evident in the work of Suzanne Briet, who famously suggested that a *documented* antelope in a zoo, as opposed to one in the wild, constitutes a document (Briet 2006 [1951]). The scope of what constitutes a document thus can be almost infinitely broadened. The criterion lies not so much in the nature of the material being stored, but in the ways in which it is treated and processed, in the differences between archival and information techniques.

Otlet, he writes about "knowledge". For various understandings (and the history) of information, see Buckland 1991, Nunberg 1996, and Day 2008. Regarding definitions of knowledge/information within the field of information science, see Stock 2015, 23-47.

Whereas the traditional archive by all means guards the integrity of fonds, the knowledge repository tends more towards atomization, breaking down material into discrete units that can then be combined and recombined[33]. While for an archivist, a single bone has nothing to say when detached from the skeleton to which it belonged, for an information manager, the bone finds its voice when indexed, classified, grouped and (re)combined with other bones sharing various features. So, for instance, Otlet proposed to extract pieces of knowledge from books and arrange these pieces on index cards, which could then be placed in different order and put together in different combinations:

> The ideal would be to strip each article or each chapter in a book of whatever is matter of fine language or repetition or padding and to collect separately on cards whatever is new and adds to knowledge (Otlet 1990 [1892], 17).

While the archive preserves the original order of documents, LIS instrumentalizes and problematizes the very ordering, exploring its productive or "generative" potential. Unlike archivists, information managers are called upon to intervene, to classify, (re)arrange and (re)organize information through the use of information technologies. Where the traditional archive insists on local principles of preservation (each fonds as a unique organism reflecting the respective administrative body), LIS elaborates and envisions universal systems, standards and ways of classifying information.

The archive is driven by living-being metaphors, it preserves what is seen as natural and organic. The knowledge repository engages more with technological and machine ones, elaborating *mechanisms*[34] and material technologies for organizing and generating knowledge. Information technologies range from reference books (Blair 2011), *fiches érudites* (Bert 2019), paper slips and card catalogues (Krajewski 2011), classifications, indexes (Day 2014, Duncan 2022) and filing cabinets (Robertson 2021) to databases, automated knowledge organization systems and the Semantic Web. It is through these information systems that an object takes on the status of a document. A document in this paradigm is what can be placed in a valid knowledge architecture, what can be indexed through operational indexing systems and classified through operational classification schemas.

All these technologies and "little tools of knowledge" (Becker, Clark 2001) are designed to ensure the retrievability of knowledge. As stated in the first of the five laws of library science formulated by S. R. Ranganathan in 1931, "Books are for use". This emphasis

---

33 Moreover, as shown by Artur Perret (2022), the foundational (epistemological) principles underlying Otlet's bibliographic system is very similar to those of the hypertext.

34 Paul Otlet, notably, claimed that "The book is a mechanism" (Otlet 1934, 422).

on the user is yet another difference between the information and archival paradigms. In the traditional archive, a user cannot do without an archivist, a necessary intermediary familiar with the structure of a particular organization and its fonds. In the knowledge repository, the organization of information is designed for the user to locate it[35].

Yet, even more importantly, in contrast to the archive, the knowledge repository has a generative function. The information technologies and knowledge systems are meant not just to store and retrieve documents but to actively contribute to the creation of new knowledge[36]. While the archive preserves sediments of administrative activities, valuing preservation as an end in itself, "knowledge machines" are tasked with producing new knowledge through the combination and recombination of existing "bits".

1.1.3. Two paradigms and the digital

The two ideal-typical models, the traditional archive and the knowledge repository, appear as two different processing modes, setting the stage for the interpretation of documents. Summarizing the points discussed above, the differences between the two models are reflected in the table:

Traditional archive	Knowledge repository
paper documents	a variety of media
stable, authentic and unique records	mobile and atomistic pieces of information/knowledge
document contextualized as a part of the whole	document contextualized through information technologies
provenance, original order, physical custody, appraisal	indexing and retrieval systems

35 Since the 1950s, the discipline of information retrieval has been developing, focusing on and thematizing "a finding or discovery process with respect to stored information" (Mooers 1951, 25). Within the realm of information retrieval, it is assumed that the knowledge organization system directly influences how a user can extract information by querying the system and to what extent their information needs can be met. On the history of information retrieval and testing of different knowledge organization technologies from indexing systems to hypertext see Ellis 1996.

36 This idea of producing knowledge through technology found its full realization in the 1980s, when the discipline of Knowledge Representation (KR) took shape. KR focuses on how knowledge can be represented and generated in automated systems. This involves defining data models, ontologies, and languages that structure and categorize knowledge in ways that are both meaningful and accessible to the machines.

focus on the donor	focus on the user
aimed at reflecting the order of the past	aimed at generating knowledge

Digital technologies and media bring their own affordances and lead to alternative configurations of storage, forms of processing documents and giving them meaning. In terms of the two paradigms outlined above, it is fair to claim that the digital turn disrupts the configurations of the traditional archive, aligning digital collections more closely with the knowledge repository model.

To start with, the very material conditions of production of documents and records has changed radically. The *Dutch manual* was originally grounded in the bureaucratic practices and technologies of the 19th century. It operated in the era when record production was paper-based and conformed to administrative and structural divisions. The outcome of these practices was a paper document, a tangible, stable and fixed record occupying a definitive place within the organization's paper workflow.

In the digital realm, archival institutions are compelled to adapt to new forms of electronic corporate information management (Cook 1994, Megill 1996). Under decentralized and fluid record-keeping systems, it is no longer possible to track document transactions that were once clearly recorded on paper. As noted by Kenneth Megill, «in the electronic age, 'document' is becoming a verb" (Megill 1996, 25), "a response to a query", a "pure function" (ibid., 27). This implies that a document is not simply produced and then stored; rather, it is actively created and recreated in response to user interaction.

Unstable and unfixed[37], the digital document no longer has a proper place of its own. Connected through a network of hyperlinks, it can be accessed in any order and in any sequence, independent of its storage location. The very idea of wholeness, so diligently safeguarded by the traditional archive (and still advocated by some contemporary archival standards), is no longer applicable or operational in the digital realm. The traditional archive seeks to preserve things in their integrity and original form; the digital archive operates not with things in their entirety but with discrete bits and pieces, with zeros and ones, devoid of any semantics per se, yet open to manipulation (Bachimont 2020).

37 For a more nuanced discussion on the stability of paper and digital documents, see Levy 1994.

As such, digital documents are function of software and hardware operations. They become accessible and interpretable only through the mediation of technologies. In the paradigm of the traditional archive, a document acquires its meaning within the context of the archival fonds. By contrast, for a digital record, the primary context is the many layers of software, without which it cannot be displayed or understood (Allen-Robertson 2018)[38].

These material and technological affordances[39] of the digital record render it essentially *unarchivable* (in a traditional archival sense). Storing a document in its original physical form has become meaningless, given that its representation is always mediated by technologies. It is no longer possible to situate and contextualize a document within a whole, as the very singular, unified whole has ceased to exist in the digital realm. The very nature of digital production negates the concept of a fixed original order. Instead, it embraces a continual process of data reconfiguration and recombination.

That said, digital documents, by virtue of their inherent media nature, align nearly perfectly with the characteristics of the document in the information paradigm: they become mobile, divisible into parts, and interconnected in various ways through hyperlinks. Like knowledge repository documents, they are contextualized through information technologies: indexing and retrieval systems, links and cross-references [40].

These alterations were recognized quite early by archivists and engendered extensive debates and upheavals in the field of archival science**.** Since the early 1980s, archivists have begun discussing the onset of the "postcustodial era," which calls for a reassessment of old archival principles (Ham 1981). This concept marked a departure from the traditional custodial approach, where archives are responsible for the physical storage of records. In the post-custodial paradigm, records remain with their creators, while archives perform support and management functions. Physical custody no longer remains the primary and defining task of an archivist. Consequently, archivists are no longer primarily seen as custodians; instead, their role shifts more towards (information) management.

Throughout the 1980s and 1990s, numerous concepts and projects aimed at "reinventing archives" (Bearman, Hedstrom 1993) emerged. This call spanned almost every

---

[38] This discussion will also continue in the next chapter dealing with the OAIS model.

[39]The affordances and limitations of digital technologies and media have been described in an extraordinary body of literature. For some informative discussions see, e.g., Levy 2011, Kirschenbaum 2012, Hayles 2012.

[40] The description of the digital production / archive is, of course, also ideal typical. In practice, paper-based practices have not disappeared, and the concept of a paperless office has not fully materialized. Yet, the emergence of this hybrid mode was significant enough to prompt archivists to acknowledge a profound disciplinary crisis as early as the 1980s.

aspect of archival theory, from the principle of provenance to the procedures of archival appraisal (Cook 1994). While a detailed examination of these debates[41] is beyond the scope of this discussion, it is noteworthy that, overall, the reevaluation of concepts generally shifted archival theory closer to information theory. So, for example, in a series of articles David Bearman offers to reformulate the principle of provenance in terms of information retrieval and proposes his model of provenance-based retrieval (Bearman, Lytle 1985). Since the 2000s, there have been several attempts to transpose archival postulates and concepts into the digital realm. For instance, the genre of digital finding aids has emerged (Trace, Dillon 2012), the principle of provenance has been formalized in the form of the PROV ontology (Lebo et al. 2013), and practices for certifying the authenticity of digital objects have been developed (Duranti, Rogers 2018).

As a result, the vocabulary of archival science has undergone significant transformation, incorporating many concepts traditionally associated with the field of Library and Information Science. For instance, the recent "Handbook of Archival Practice" (Franks 2021) includes, alongside traditional entries such as "accessioning," or "appraisal," terms like "taxonomy" and "linked data," "digital curation" and "digital forensics", "information analysis" and "information governance".

Another aspect that merits at least a brief mention is the changing role of the archivist. For one thing, this change is driven by the availability of digital technologies for data collection and storage. The tasks of collecting, preserving, and archiving are no longer exclusive to the professional archivist. Therefore, as Dutch archivist Eric Ketelaar has proclaimed, in the digital era, "everyone is an archivist". (Ketelaar 2006). This claim is exemplified by the appearance of "participatory" and "community" archives, where diverse groups gather materials for their own communal memories, thereby blurring the lines between the roles of the user and the archivist.

For another thing, the convergence of archivistics with information science has prompted a reassessment and questioning of the archivist's role within the profession itself. From the 1980s onwards, there has been an increasing concern about maintaining the distinct identity of archivists in the digital world and clearly differentiating their roles from those of information managers. A somewhat common response from the archival community has been a call for archivists to adopt a more proactive role. As influential Canadian archivist Terry

[41] This debate is summarized by Cunnigham 2010. For a detailed analysis of different national archival traditions in relation to the digital turn, see Couture, Lajeunesse,2014.

Cook suggests, archivists should “stop being custodians of things” (Cook 1994, 304) and instead, “guide our sponsors and users from masses of specific information on to knowledge” (ibid., 306). Quite remarkably, the role of the archivist is reinterpreted here as that of a guide to *knowledge*, in contrast to the information manager who focuses on handling *information*. Consequently, the archive’s mission is redefined as not just preserving documents, but safeguarding knowledge itself.

The digital archive challenges many if not all of the “immutable facts” of archival science, leading to an increasingly blurred definition of what constitutes an archive. Before the digital came into play, what was considered an archive strictly adhered to the criteria set by archival science. Yet, the “reinvented” (digital) archive scarcely aligns with these standards.

Situating the digital archive somewhat between the traditional archive and the knowledge repository seems to be a fruitful way to articulate some of its tensions. On the one hand, digital technologies and media are steering the archive towards the “hidden tradition” of the knowledge repository. While the traditional archival principles need to be reformatted, rearticulated, and reinvented for the digital archive, the techniques of information science are quite “naturally” applicable to it. In the digital realm, the archive finds itself dealing with terms and techniques initially alien to it, such as *knowledge*, *user*, *representation* and *information technologies*.

On the other hand, despite the infrastructural transition, the digital archive evidently still retains the “genre memory” of the traditional archive. We have witnessed this tendency to assess digital archives through the lens of traditional standards. This is evident in attempts to reinvent traditional principles as applied to the digital realm. It is also apparent in how archivists refuse to recognize many digital collections as archives since they do not conform to the *Ius Archivi*. The very concepts of digital preservation and digital authenticity clearly originate from the configurations of the traditional archive. This perspective is present not only within the archival profession, but also at the level of user expectations and at the level of practices, which are in many ways still influenced by the traditional archival paradigm.

However, in the new *standards* of (digital) preservation, like the one we will examine in the next chapter, there is no trace left of traditional principles. Instead, fundamentally new techniques, conceptions of the archive and preservation, and even an entirely new vocabulary are introduced.

1.2. The OAIS archive: pragmatic and hermeneutical turns

> Standards are the distilled wisdom of people with expertise in their subject matter.
> International Organization for Standardization (ISO)[42]

The OAIS Reference Model is to digital preservation what the *Dutch manual* is to traditional archiving. Often referred to as the *lingua franca* of digital preservation, it provides and prescribes a common framework for discourse on the digital archive. The standard does not dictate specific actions or implementations. Instead, it sets the very language of the discussion and prescribes the basic terms and definitions. As such, the model exemplifies what sociologist Laurent Thévenot referred to as "the investment in forms" – a form-giving activity, that establishes and stabilizes shared frames of evaluation and valorization (Thévenot 1984).

This chapter explores how the OAIS standard *gives form* to the digital archive: how and in what terms does the OAIS model define the archive and the process of archiving? What, how, and for whom does the archive preserve, according to OAIS? How does the OAIS archive align with the archival and information paradigms outlined in the previous chapter?

1.2.1. The OAIS model: an overview

Interestingly, the principal standard for digital archiving emerged neither from archivists, nor from historians, but from (computer) scientists. The Open Archival Information System (OAIS) model was developed by the Consultative Committee for Space Data Systems (CCSDS), an organization typically engaged in standardizing data from space missions. Its primary task is to ensure interoperability among various national space agencies. To achieve this, the committee focuses on developing data and system standards, in collaboration with the International Organization for Standardization (ISO).

The development of the OAIS model spanned almost a decade, from the initial proposal to ISO in April 1994 to its adoption as an ISO standard in 2002. The original idea of establishing a universal archival system materialized in the early 1990s. At that time, various independent initiatives focusing on digital preservation were already in place, yet there lacked a unified standard for digital archiving. The development of the standard was led by a small

42 [https://www.iso.org/standards.html]

team of programmers and engineers under the guidance of Don Sawyer, a computer scientist and the head of the Office of Standards and Technology at NASA.

Christopher Lee provides an in-depth description of the negotiations over the reference model in his dissertation (2005)[43]. As chronicled by Lee, a crucial aspect of developing the model was the wide-ranging discussion and participation from diverse groups and organizations. The participation extended beyond the space data community to include government agencies, private sector companies, and academic institutions. The standard underwent numerous revisions, taking into account a wide array of formal and informal reviews and feedback received from various stakeholders. Throughout the development phase, dozens of workshops and public talks on the Reference Model took place. According to Lee, it is precisely due to this "unique meeting and decision-making structure" (Lee 2009, 4022) that the OAIS model has become a universal standard, extending beyond the preservation of space data.

That said, while the OAIS model seeks to offer a universal archive framework, as we will see, it is inspired by scientific data preservation. Just as traditional archives built their entire system around the case of administrative records, so too does the OAIS model center its primary use-case on the preservation of (big) scientific data. In this sense, the genealogy of the standard – its birth "out of the spirit of science" – is instructive.

The process of drafting and refining the model was iterative, involving repeated cycles of review and improvement. Draft versions of the reference model were released for feedback in 1997 and 1999. By June 2000, it had progressed to being published as a draft ISO Standard. After a final phase of review and revisions, it gained approval and was officially published in 2003. In 2012, a new revised version was published. This update, in particular, took into account some feedback from archivists. A key addition was the definition of "authenticity" – a concept that was notably missing in the original version of the standard.

The latest draft of the standard was released by the CCSDS in 2019. This is a 151-page document that delineates the key terms and principles of digital long-term preservation. Currently, it is in the process of being ratified to become a new ISO standard[44].

The OAIS Model has become a "seminal document" in the field of digital preservation. It has been adopted and widely discussed within the archival community (O'Meara 2016, Franks 2021). As summarized in one of the model reviews,

---

43 See also Giaretta 2011, a book in which one of the developers of the standard details the intricacies of the model and offers several use-cases.

44 Since this version has already been approved as "recommended practices", I will use it for quotations. It is also available online [https://public.ccsds.org/Lists/CCSDS%206500P21/650x0021.pdf]

> The most important achievement of the OAIS reference model to date is that it has become almost universally accepted as the *lingua franca* of digital preservation, shaping and sustaining conversations about digital preservation across disparate domains, and supplying a general mapping of the landscape that stewards of our digital heritage must navigate in order to secure the long-term availability of digital materials (Lavoie 2014, 3).

Yet beyond language, OAIS is also deeply incorporated in archival practice. Major national archival actors, including the Library of Congress, British Library, Bibliothèque nationale de France (BnF), and U.S. National Archives, built their digital repositories in accordance with OAIS recommendations. The model also serves as the basis for other broadly accepted standards in digital preservation, including the "Producer-Archive Interface Standard" (PAIS), the "Producer-archive Interface Methodology Abstract Standard" (PAIMAS), and the PREMIS Data Dictionary for Preservation Metadata. OAIS also underlies the Trusted Digital Repository (TDR) Checklist, an audit system that qualifies an archive as being "trustworthy". Last but not least, a great deal of archival technologies, including repository platforms and content management systems (such as *DSpace* or *Fedora*) conform to the standard. The OAIS model thus does not simply offer a common discourse about the archive, but its concepts and terms are deeply embedded in the archival technologies themselves.

1.2.2. Toward the definition of archive and archiving

The OAIS model puts forward a new language of archiving, explicitly distinguishing it from the vocabulary of traditional archival science. The standard specifies that "the approach is taken to use terms that are not already overloaded with meaning so as to reduce conveying unintended meanings" (CCSDS Pink Book 2019, 1-7). For example, what is traditionally known as a "record" in archival terms is approximately equivalent to the "Content Information within an Archival Information Package" in the OAIS model (ibidem).

OAIS's definition of archive, though, is quite straightforward and lapidary: it is defined as "an organization that intends to preserve information for access and use by a Designated Community" (ibidem). The first thing to notice in this definition is the way in which the object of archiving is defined. The OAIS archive preserves not records, documents, or heritage, but *information* seen as "any type of knowledge that can be exchanged" (ibid., 1-11). Relying on the concept of information, the definition makes no difference between objects of storage: the OAIS archive can preserve any type and genre of items, whether born-digital, digitized, or non-digital, as long as they are qualified as being *information.*

Information can be of any type; but what defines it is its pragmatics – the fact that it is (re)usable and understandable. The first and crucial objective of the OAIS archive, evident in its very definition, is maintaining the use-value, intelligibility, and usability of the information for the users. The digital archive is thus intended not only and not so much to preserve the items themselves, but to ensure that they remain comprehensible and usable for the target audience in the long run.

In this pragmatic redefinition of the archive, two concepts are of particular importance: Designated Community and the long-term. The Designated Community is an (imagined) target community of users for a specific information set,

> an identified group of potential Customers who should be able to understand a particular set of information in ways exemplified by the Preservation Objectives. The Designated Community may be composed out of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time (ibid., 1-10).

Whereas the traditional archive mainly articulates its relationship with the "producer" (donor/administrative body), the digital archive rather brings to the fore its relationship with the user. The OAIS Archive does not just keep information for its own sake, it always does so with an eye toward a particular Designated Community sharing a common Knowledge Base[45], which enables its members to comprehend given information.

Notably, according to the OAIS model, it is the archive itself that is responsible for selecting and defining its potential users. Explicitly defining the 'Designated Community' is a fundamental task stated in the compliance rules of the OAIS model. In other words, an OAIS-compliant archive cannot be considered as such without specifying the designated community it serves.

Whereas the traditional archive focuses on selecting what to preserve, the digital archive emphasizes the selection of whom to preserve for. While this new pragmatic concept of the archive has become the universal standard of the digital preservation, it has raised a good deal of challenges for archivists (McMeekin 2011). Not coincidentally, it is the notion of Designated Community that caused particular difficulties, being "a source of frustration to some in digital preservation, particularly librarians who find the need to specify a particular community to be at odds with their professional, and sometimes legal, mandate to serve broad populations" (Bettivia 2016, 1).

---

45 Knowledge Base is formally defined as "a set of information, incorporated by a person or their proxy system that allows that person or their proxy system to understand the received information" (ibid., 1-11).

The imperative to select the groups of users and to target their interests is certainly in conflict with the traditional archival principles. The traditional archive is concerned with a "truthful" reflection of the past, while the digital archive aims for future use. Whereas the traditional archive is focused on the donor, the digital one is centered on the user. The traditional archive preserves records in order to (*impartially* and *objectively*) reflect the life of the administrative body. The OAIS archive preserves information in a deliberately partial manner, so that it can be used by a particular group of users.

Furthermore, the digital archive does not store information for eternity, it operates in the long-term, understood as

> a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community or changes to the Designated Community's Knowledge Base, on the information being held in an OAIS. This period extends into the indefinite future (CCSDS Pink Book 2019, 1-11–1-12).

The concept of the long-term addresses two types of change: changing technologies (obsolescence of formats, media, software and hardware) and changing knowledge of designated communities. In this context of changing technological and epistemological horizons, the archive is called upon to ensure the comprehensibility of stored information passed from producer to user.

To this end, the concept of *information packages* is introduced (as opposed to the *fonds* or *record groups* of the traditional archive). Three types of such packages (Fig. 1.1) are distinguished: Submission Information Package (SIP), Archival Information Package (AIP), and Dissemination Information Package (DIP).

SIP refers to the way in which content is submitted to the archive by the "producer". SIPs are then converted and remodeled into Archival Packages (AIPs), in which the content will be preserved within the archive. Being the responsibility of the OAIS archive, this conversion can be performed in different ways: the archive may, for example, aggregate multiple SIPs into a single AIP or, conversely, many SIPs from different sources may be "unbundled and recombined in different ways to produce many AIPs" (ibid., 2-8). The OAIS archive thus neither preserves nor even thematizes the "original order". In violation of the core principles of the traditional archive and following the "information paradigm", it stands free to recombine and reassemble its packages.

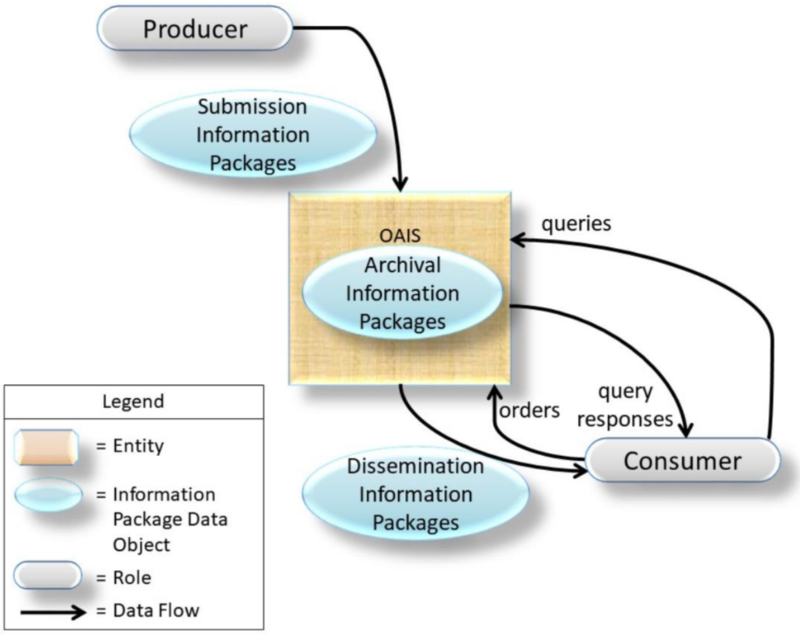


Figure 1.1. OAIS Information Packages (ibid., 2-10).

The third type of information package in OAIS is the Dissemination Package, i.e., what the archive provides to the user in response to their request. In turn, it can be assembled from one or more AIPs, depending on the user's query[46]. SIP – what the archive receives as input – and DIP – what it outputs to the user – thus stand apart from each other and constitute two different sets of "information". The Archive Package then serves as an intermediate link between them, providing the transmission (if not *translation*) of the received information into the information demanded by the user.

The process of preservation thus appears as an (ongoing) transformation of data aimed at maintaining its usability and intelligibility. The OAIS archive does not imply the idea of permanence; instead, the concept of change is integral to its very essence. The preservation is no longer seen as the *conservation of the past* but rather as a *transmission to the future.*

1.2.3. OAIS information model and the concept of understanding

The OAIS information packages store *information objects*, made up of *data objects* and their *representation information*. *Data object* is the content to be preserved, either a physical or digital entity. *Representation Information* is what "allows for the full interpretation of the data

[46] Different "functional entities" within the OAIS archive are responsible for transformations and inspections of archival packages. Each of these entities and their functions are detailed in the Standard (4-1–4-22). Also see the detailed discussion in Giaretta 2011.

into meaningful information" (ibid., 4-22–4-23), what makes it possible to read and understand the data object (Fig. 1.2)[47].

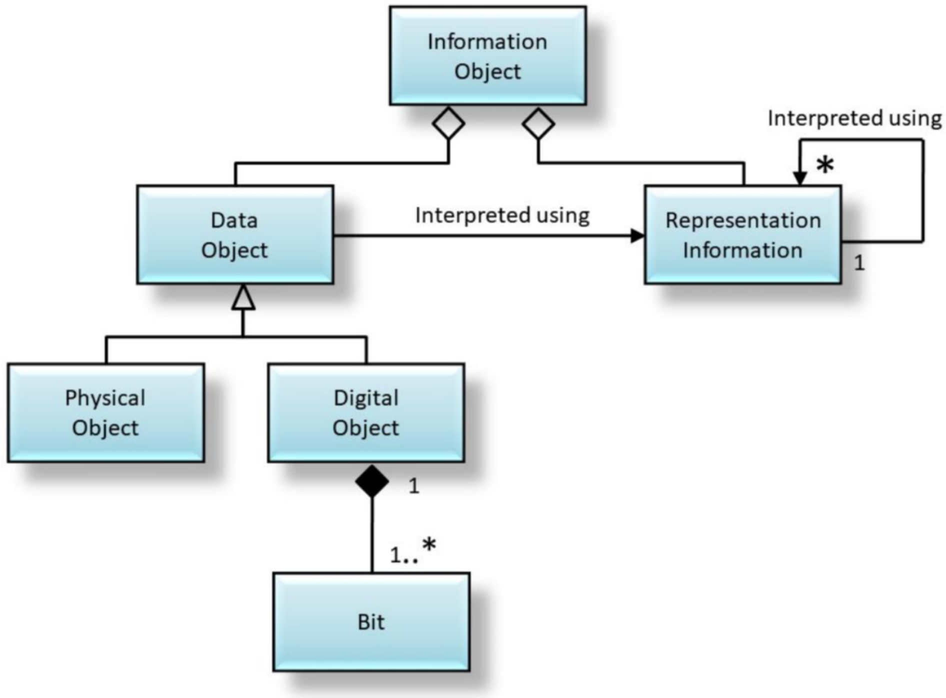


Figure 1.2. OAIS Information Object (ibid., 4-23)

In the configuration of a traditional archive, an archival document is what is kept in its proper place. The digital archive constitutes archival documents through providing them with suitable *Representation Information*, i.e. some instructions of how they are to be read. It is through this *Representational Information* that an arbitrary digital object acquires the status of archival document that is preserved and cared for.

Representation information is categorized into three types: *Structure*, *Semantic* and *Other*. Structure information describes formats and types of data, so that they can be interpreted as numbers, arrays, images, tables. Semantic information explains the meaning of data and the way they are related to each other. Finally, the third type, referred to as Other Representation Information, describes any other information needed to interpret the data that is not covered by the first two types[48].

---

47 In addition to the *Representation Information*, the *Information Object* may also include other types of description, collectively known as Preservation Description Information (PDI). PDI, in turn, includes Reference Information, Context Information, Provenance Information, Fixity Information, Access Right Information (ibid., 4-33—4-34).

48 This type was introduced later than the others (in the 2012 version of the standard), suggesting that the two types of information in practice were insufficient.

To give the OAIS's own example, suppose that the data object to be preserved is a PDF document containing some text in the English language. To secure its intelligibility in the long term, the archive has to ensure that the document format remains readable and accessible, the necessary technologies for interaction with it are available, and the document's content stays accessible. As part of the structure representation information, it would need to preserve the standard of the PDF format. For semantic information, it might require explanations of terms used in the text, or potentially, a dictionary and grammar guide for the English language. Additionally, it would need to preserve software capable of reading PDF files, which would be categorized as Other Representation Information (ibid., 4-25).

Even in this simplest example, where the object of preservation is a text encoded in a universal and standard format, the archive is meant to store and collect all those materials that render it readable. The *Representation Information* in this case will certainly exceed the size of the data object itself. Not to mention the fact that it is fluid and has to be updated, refined and enriched in accordance with the changing technologies and Knowledge Base of the Designated Community. Further still, OAIS keeps in mind that in the long-term users may also face problems in interpreting the very *Representation information*. The standard therefore stipulates that *Representation information* should also contain all the information necessary for its own interpretation. At its extreme, OAIS sets in motion a kind of infinite loop aimed at ensuring readability of all its components. Archiving, as described in the OAIS, thus implies a tireless hermeneutic endeavor: the archive at all times is supposed to question and assure its intelligibility.

At its core, the standard emphasizes that preserving objects (data or content) alone is not enough; in order to ensure their readability, the archive must also maintain the interpretive framework within which these objects are understood. Information is thus seen not merely as content, as "a noble substance that is indifferent to the transformation of its vehicles" (Nunberg, 1996, 107), but as content *in relation to* or *as function of* the means (or context) of its interpretation.

As Bruno Bachimont observes, by adopting this definition of information, OAIS follows the very nature of the digital:

> The OAIS information model takes full account of the characteristics of digital data, notably the fact that it is "anonymous" and that its meaning is always the result of arbitrary conventions applied to it. In this respect, this standard is profoundly original and has no equivalent (Bachimont 2020, 206; my translation).

The (digital) object, in other words, is never given to us *per se*. Even the most "innocent" and self-evident item requires some interpretive framework, which in the long term (under conditions of technical and epistemological volatility) is subject to change. Whereas in the traditional archive things were meant to speak for themselves, in the OAIS archive they can only speak from within a particular technological and hermeneutical horizon.

For the humanities, this idea is hardly new, though the horizon of understanding in it is maintained not through Representation Information, but rather via *tradition*. What is surprising, however, is that this *topos* of understanding – the "classical" method of the humanities – penetrates the science-inspired standard and turns out to be its leitmotif. Essentially, the OAIS model interprets the archive hermeneutically: the mission of the archive extends beyond merely taking objects under its care to ensuring their understanding. The archivist's role has evolved from simply selecting what will be preserved to having a more explicit control over the meaning and interpretation of these items. It is the archivist who gets to decide "what parts of the Content Information are the Content Data Object and what parts are the Representation Information" (CCSDS Pink Book 2019, 4-30). They get to choose the *Representation information* for archival items, and thus set the horizon for their interpretation.

That said, turning *understanding* into a key archival concept, the OAIS model adopts a rather mechanistic view of it. Understanding is seen as almost a technical operation:

> As a practical matter, the OAIS needs to have enough Representation Information associated with the bits of the Content Data Object in the Content Information that it feels confident that the members of the Designated Community can enter the Representation Information Network with enough knowledge to begin accurately interpreting the Representation Information (ibid., 4-29).

According to the OAIS, one can achieve an "independent" and "accurate" understanding if there is a sufficient amount of Representation Information available. To understand a text "independently" and "accurately," one needs to know the language in which it is written and the rules of grammar; to decipher a table, one must be familiar with its encoding, and so on. In the standard, "understandability" is almost equated with "usability"[49]. Take, for example, one of the few OAIS humanities examples – the preservation of a piece of music. As the standard suggests, the "preservation would be regarded as successful if the musical work can be reperformed in future in a way which the composer would regard as the same" (ibid., 4-27).

[49] For some examples see Giaretta 2011, 169.

“Understandability” is viewed more as scientific reproducibility than as “being open for interpretation”.

At this point, one envisions certain limits of the standard. “Understanding” as a technical concept may work for data replicability, but is hardly operational in the context of history and humanities in general. What does it mean to make the residues of science accessible to historical understanding? What exactly needs to be preserved to maintain an interpretive horizon for an obsolete scientific instrument, a piece of paper, or a piece of code? Is technical documentation and format information sufficient, or is a broader (historical) context necessary? These are the kinds of questions that need to be addressed in relation to a wide array of historical artifacts, if the archive is viewed as a conduit of understanding (these are the questions that the subsequent chapters explore).

The OAIS model exemplifies the emergence of fundamentally new, hybrid practices in conceptualizing and discussing the digital archive. Or, to put it another way, in/through it, the digital archive is taking on forms that are radically different from those of the traditional archive. Perhaps most importantly, the standard introduces into archival discourse the questions of understanding, use, and production of information/knowledge. The archive no longer stores documents for their own sake, but stores “information” to be understood, (re)used, and to produce new information.

1.3. History of science and its archive

Between archival and information studies, I have not yet considered a third major actor of the field: history (of science)[50]. The archive is taking shape not only through the practices of archivists and information managers, but also through the work of historians.

Archivists and historians often disagree on what exactly should be preserved for the sake of the future. Chronicling the dynamics of their relationship, Francis Blouin and William Rosenberg point to what they call *the archival divide* – "a deeply conceptual separation based on different readings of the relation between past and present, on how pasts can and should be literally and figuratively processed" (Blouin, Rosenberg 2011, 93). As I will show, it is the historian's "readings" that shape the idea of the scientific residue and the practices of its archiving and interpretation.

In this chapter, I present three episodes of the interaction between the history of science and its archive. Each of these three projects imagines and assembles the archive of science in its own way. Yet, taken together, they contribute to a cohesive story of how the archive of science has been conceptualized and nurtured by historians.

1.3.1. Assembling the archive of "living science"

The question of how to archive and document recent science came into discussion in the late 1950s[51]. In the United States[52], the issue was brought into focus at the Conference on Science Manuscripts held in Washington D.C. in May 1960 (Anderson 2013). By this point, America had only one archive of a twentieth-century scientist, that of Enrico Fermi (Hamer 1961). The overall sense was spreading that "time was running out" and the "rich heritage" of "the heroic age of American science" was "slipping through fingers", as the historian Gerald J. Gruman put it in 1958 (Gruman 1958, 1471). "Preserving the stuff of history" (to quote the title of Gruman's article) suddenly turned out to be an urgent task that historians and archivists had to tackle together.

Beginning in the 1960s in America and subsequently in Europe, a number of initiatives and projects were launched to address this issue. One of the primary responses was the establishment of "discipline history centers" (Warnow-Blewett 1992). These are permanent institutions, typically affiliated with a science center or university, tasked with documenting

[50] The forth major actor will be the scientist, whose memorial practices delineate and shape the archive.
[51] For histories of scientific collections prior to this, cf. Hunter 1998, Keller et al. 2018.
[52] On other national contexts see e.g. MacLeod, Friday 1973, Charmasson 2006.

and archiving a particular discipline. The first such center was the AIP Center for History of Physics, opened in 1960. Following the AIP, similar centers focusing on a variety of disciplines were established, including the Charles Babbage Institute for the History of Information Processing (University of Minnesota), the Archives of History of American Psychology (University of Akron) and the Beckman Center for the History of Chemistry. By the early 1990s, there were already eighteen such centers in the United States (ibid., 285-290). Similar documentation centers for recent science have been established in many countries. Some of the examples include the Contemporary Scientific Archives Centre in Britain (founded in 1973), the Australian Science Archives Project (1980s), the Centre de Recherche en Histoire des Sciences et des Techniques in France (1986).

In parallel, from the early 1960s onwards, new archival services were established in many institutions associated with science and engineering, including Stanford University (1965), California Institute of Technology (1968), MIT (1977), and CERN (1979).

Alongside the establishment of the discipline history centers, documentary projects were initiated to collect materials for the history of specific disciplines. The first of such initiatives was launched in 1961, when the American Physical Society and American Philosophical Society jointly proposed a project to collect materials on the history of quantum physics. Funded by a grant from the National Science Foundation, the project spanned from 1961 to 1964 and was headed by none other than Thomas Kuhn, who had just finalized the bulk of *The Structure of Scientific Revolutions*.

The project was spurred by the same growing sense that great scientists and their legacies were vanishing before one's eyes. As physicist John Archibald Wheeler put it in his foreword to the final report of the project, "the bell has been tolling; time is short. Einstein died in 1955, von Neumann in 1957, Pauli in 1959, and Schroedinger in 1961" (Kuhn et al. 1967, vi). None of them had ever been interviewed at length by a historian. At the time, their manuscripts, materials, and documents remained in private possession, with some being lost and others dispersed across numerous scientific institutions around the world. As a result, by the early 1960s, as Wheeler noted, there was less material available on the history of quantum physics than that on the physics of the 18th century. This disparity in available resources made it more feasible for a historian to write about the history of Newtonian physics than about the physics of Einstein. Recent science was seen as a "disappearing object" to be collected, classified and preserved. Therefore, the mission of the project was framed as follows:

> To remedy this deficiency in source materials and to open the doors to a new understanding of how science operates, it was agreed that the project should seek to ensure the preservation of whatever

manuscript materials remain; to record recollections and commentary by the living participants and their close associates; and to prepare for publication a catalog of this material (ibid., viii).

The primary *raison d'être* of the project was collecting oral histories of "big" physicists involved in developing quantum theory[53]. The selection of the respondents was conducted through biographical and bibliographical searches. The project commenced with the compilation of 170 biographies "of men who contributed significantly to the development of quantum physics" (ibid., 2). Subsequently, 175 interviews were conducted with nearly a hundred physicists, which were meant to shed light on the "major episode in the development of modern science" (ibid., 1)[54].

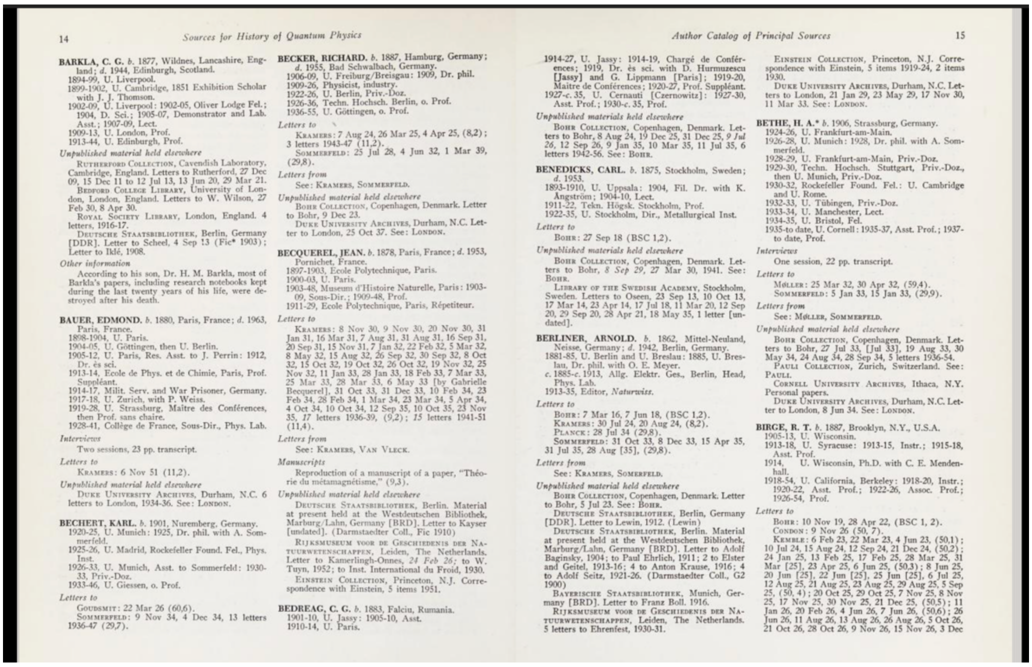
In addition to the interviews, the project also collected scientific residues of the chosen physicists: drafts, letters, laboratory notebooks, photographs. In order to localize the materials, letters and questionnaires were sent to scientific journals and institutions across Europe and the US, as well as to the selected physicists.

The selection of documents was driven by two considerations: the significance of scientists and the disciplinary focus. First, the project stressed the importance of individual contributions for the history of quantum physics. It identified "the most significant" collections available: the report lists those of Niels Bohr (at his institute in Copenhagen), Albert Einstein (Institute for Advanced Study), Wolfgang Pauli (at the time still in his wife's possession, later at CERN), O. W. Richardson (University of Texas), and the collections at the American Institute of Physics (AIP). The resulting archive therefore presents the history of the discipline as the sum of personal histories. The final report indexing the collected sources is constructed as an author's catalog (Fig. 1.3), with the name of the scientist serving as the main reference point. Second, those materials were selected that concerned the quantum theory: correspondence, drafts of papers on quantum physics, research notebooks and "a representative selection of lectures on topics relating to the quantum" (ibid., 7). Microfilm copies were made of all the selected materials, resulting in more than 100 rolls of films, which together with the interview tapes formed the Archives for the History of Quantum Physics (AHQP)[55].

In her detailed analysis of Kuhn's project, historian of science Anke te Heesen finds its groundbreaking nature in the appeal to oral history. In her view, it is the use of the spoken word as a new source that marks a paradigm shift to the new (post-Kuhnian) way of doing history of

[53] Among them was Niels Bohr, who notably gave his last interview to Kuhn a day before his death.
[54] For more on the project and Kuhn's disillusionment in the interview see Heesen 2020 and 2022.
[55] The archives is housed in the three repositories: Library of the American Philosophical Society, Library of the University of California, and Bohr archive in Copenhagen.

science. According to Heesen, the project represents one of the first attempts to narrate the history of science not from the perspective of its achievements and ideas, but to explore how these ideas and achievements (as well as failures) are produced. Recording life histories and articulating living memories of physicists, the project draws attention to the "human, emotional side of scientific work" (Heesen, 2020, 88), gives voice to scientists themselves, and addresses science in the making as opposed to the history of ideas.



14 *Sources for History of Quantum Physics*

BARKLA, C. G. *b.* 1877, Wildnes, Lancashire, England; *d.* 1944, Edinburgh, Scotland.
1894-99, U. Liverpool.
1899-1902, U. Cambridge, 1851 Exhibition Scholar with J. J. Thomson.
1902-09, U. Liverpool: 1902-05, Oliver Lodge Fel.; 1904, D. Sci.; 1905-07, Demonstrator and Lab. Asst.; 1907-09, Lect.
1909-13, U. London, Prof.
1913-44, U. Edinburgh, Prof.
Unpublished material held elsewhere
Rutherford Collection, Cavendish Laboratory, Cambridge, England. Letters to Rutherford, 27 Dec 09, 15 Dec 11 to 12 Jul 13, 13 Jun 20, 29 Mar 21.
Bedford College Library, University of London, London, England. Letters to W. Wilson, 27 Feb 30, 8 Apr 30.
Royal Society Library, London, England. 4 letters, 1916-17.
Deutsche Staatsbibliothek, Berlin, Germany [DDR]. Letter to Scheel, 4 Sep 13 (Fic* 1903); Letter to Iklé, 1908.
Other information
According to his son, Dr. H. M. Barkla, most of Barkla's papers, including research notebooks kept during the last twenty years of his life, were destroyed after his death.

BAUER, EDMOND. *b.* 1880, Paris, France; *d.* 1963, Paris, France.
1898-1904, U. Paris.
1904-05, U. Göttingen, then U. Berlin.
1905-12, U. Paris, Res. Asst. to J. Perrin: 1912, Dr. ès sci.
1913-14, Ecole de Phys. et de Chimie, Paris, Prof. Suppléant.
1914-17, Milit. Serv. and War Prisoner, Germany.
1917-18, U. Zurich, with P. Weiss.
1919-28, U. Strassburg, Maître des Conférences, then Prof. sans chaire.
1928-41, Collège de France, Sous-Dir., Phys. Lab.
Interviews
Two sessions, 23 pp. transcript.
Letters to
Kramers: 6 Nov 51 (11,2).
Unpublished material held elsewhere
Duke University Archives, Durham, N.C. 6 letters to London, 1934-36. See: London.

BECHERT, KARL. *b.* 1901, Nuremberg, Germany.
1920-25, U. Munich: 1925, Dr. phil. with A. Sommerfeld.
1925-26, U. Madrid, Rockefeller Found. Fel., Phys. Inst.
1926-33, U. Munich, Asst. to Sommerfeld: 1930-33, Priv.-Doz.
1933-46, U. Giessen, o. Prof.
Letters to
Goudsmit: 22 Mar 26 (60,6).
Sommerfeld: 9 Nov 34, 4 Dec 34, 13 letters 1936-47 (29,7).

BECKER, RICHARD. *b.* 1887, Hamburg, Germany; *d.* 1955, Bad Schwalbach, Germany.
1906-09, U. Freiburg/Breisgau: 1909, Dr. phil.
1909-26, Physicist, industry.
1922-26, U. Berlin, Priv.-Doz.
1926-36, Techn. Hochsch. Berlin, o. Prof.
1936-55, U. Göttingen, o. Prof.
Letters to
Kramers: 7 Aug 24, 26 Mar 25, 4 Apr 25, (8,2); 3 letters 1943-47 (11,2).
Sommerfeld: 25 Jul 28, 4 Jun 32, 1 Mar 39, (29,8).
Letters from
See: Kramers, Sommerfeld.
Unpublished material held elsewhere
Bohr Collection, Copenhagen, Denmark. Letter to Bohr, 9 Dec 23.
Duke University Archives, Durham, N.C. Letter to London, 25 Oct 37. See: London.

BECQUEREL, JEAN. *b.* 1878, Paris, France; *d.* 1953, Pornichet, France.
1897-1903, Ecole Polytechnique, Paris.
1900-03, U. Paris.
1903-48, Museum d'Histoire Naturelle, Paris: 1903-09, Sous-Dir.; 1909-48, Prof.
1911-29, Ecole Polytechnique, Paris, Répetiteur.
Letters to
Kramers: 8 Nov 30, 9 Nov 30, 20 Nov 30, 31 Jan 31, 16 Mar 31, 7 Aug 31, 31 Aug 31, 16 Sep 31, 20 Sep 31, 15 Nov 31, 7 Jan 32, 22 Feb 32, 5 Mar 32, 8 May 32, 15 Aug 32, 26 Sep 32, 30 Sep 32, 8 Oct 32, 15 Oct 32, 19 Oct 32, 26 Oct 32, 19 Nov 32, 25 Nov 32, 11 Jan 33, 28 Jan 33, 18 Feb 33, 7 Mar 33, 25 Mar 33, 28 Mar 33, 6 May 33 [by Gabrielle Becquerel], 31 Oct 33, 31 Dec 33, 10 Feb 34, 23 Feb 34, 28 Feb 34, 1 Mar 34, 23 Mar 34, 5 Apr 34, 4 Oct 34, 10 Oct 34, 12 Sep 35, 10 Oct 35, 23 Nov 35, *17* letters 1936-39, (9,2); *15* letters 1941-51 (11,4).
Letters from
See: Kramers, Van Vleck.
Manuscripts
Reproduction of a manuscript of a paper, "Théorie du métamagnétisme," (9,3).
Unpublished material held elsewhere
Deutsche Staatsbibliothek, Berlin. Material at present held at the Westdeutschen Bibliothek, Marburg/Lahn, Germany [BRD]. Letter to Kayser [undated]. (Darmstaedter Coll., Fic 1910)
Rijksmuseum voor de Geschiedenis der Natuurwetenschappen, Leiden, The Netherlands. Letter to Kamerlingh-Onnes, *24 Feb 26;* to W. Tuyn, 1952; to Inst. International du Froid, 1930.
Einstein Collection, Princeton, N.J. Correspondence with Einstein, 5 items 1951.

BEDREAG, C. G. *b.* 1883, Falciu, Rumania.
1901-10, U. Jassy: 1905-10, Asst.
1910-14, U. Paris.

Author Catalog of Principal Sources 15

1914-27, U. Jassy: 1914-19, Chargé de Conférences; 1919, Dr. ès sci. with D. Hurmuzescu [Jassy] and G. Lippmann [Paris]; 1919-20, Maitre de Conférences; 1920-27, Prof. Suppléant.
1927-*c.* 35, U. Cernauti [Czernowitz]: 1927-30, Asst. Prof.; 1930-*c.* 35, Prof.
Unpublished materials held elsewhere
Bohr Collection, Copenhagen, Denmark. Letters to Bohr, 8 Aug 24, 19 Dec 25, 31 Dec 25, *9 Jul 26,* 12 Sep 26, 9 Jan 35, 10 Mar 35, 11 Jul 35, 6 letters 1942-56. See: Bohr.

BENEDICKS, CARL. *b.* 1875, Stockholm, Sweden; *d.* 1953.
1893-1910, U. Uppsala: 1904, Fil. Dr. with K. Ångström; 1904-10, Lect.
1911-22, Tekn. Högsk. Stockholm, Prof.
1922-35, U. Stockholm, Dir., Metallurgical Inst.
Letters to
Bohr: 27 Sep 18 (BSC 1,2).
Unpublished materials held elsewhere
Bohr Collection, Copenhagen, Denmark. Letters to Bohr, *8 Sep 29,* 27 Mar 30, 1941. See: Bohr.
Library of the Swedish Academy, Stockholm, Sweden. Letters to Oseen, 23 Sep 13, 10 Oct 13, 17 Mar 14, 23 Apr 14, 17 Jul 18, 11 Mar 20, 12 Sep 20, 29 Sep 20, 28 Apr 21, 18 May 35, 1 letter [undated].

BERLINER, ARNOLD. *b.* 1862, Mittel-Neuland, Neisse, Germany; *d.* 1942, Berlin, Germany.
1881-85, U. Berlin and U. Breslau: 1885, U. Breslau, Dr. phil. with O. E. Meyer.
c. 1885-*c.* 1913, Allg. Elektr. Ges., Berlin, Head, Phys. Lab.
1913-35, Editor, *Naturwiss.*
Letters to
Bohr: 7 Mar 16, 7 Jun 18, (BSC 1,2).
Kramers: 30 Jul 24, 20 Aug 24, (8,2).
Planck: 28 Jul 34 (29,8).
Sommerfeld: 31 Oct 33, 8 Dec 33, 15 Apr 35, 31 Jul 35, 28 Aug [35], (29,8).
Letters from
See: Kramers, Somerfeld.
Unpublished material held elsewhere
Bohr Collection, Copenhagen, Denmark. Letter to Bohr, 5 Jul 23. See: Bohr.
Deutsche Staatsbibliothek, Berlin, Germany [DDR]. Letter to Lewin, 1912. (Lewin)
Deutsche Staatsbibliothek, Berlin. Material at present held at the Westdeutschen Bibliothek, Marburg/Lahn, Germany [BRD]. Letter to Adolf Baginsky, 1904; to Paul Ehrlich, 1911; 2 to Elster and Geitel, 1913-16; 4 to Anton Krause, 1916; 4 to Adolf Seitz, 1921-26. (Darmstaedter Coll., G2 1900)
Bayerische Staatsbibliothek, Munich, Germany [BRD]. Letter to Franz Boll. 1916.
Rijksmuseum voor de Geschiedenis der Natuurwetenschappen, Leiden, The Netherlands. 5 letters to Ehrenfest, 1930-31.

Einstein Collection, Princeton, N.J. Correspondence with Einstein, 5 items 1919-24, 2 items 1930.
Duke University Archives, Durham, N.C. Letters to London, 21 Jan 29, 23 May 29, 17 Nov 30, 11 Mar 33. See: London.

BETHE, H. A.* *b.* 1906, Strassburg, Germany.
1924-26, U. Frankfurt-am-Main.
1926-28, U. Munich: 1928, Dr. phil. with A. Sommerfeld.
1928-29, U. Frankfurt-am-Main, Priv.-Doz.
1929-30, Techn. Hochsch. Stuttgart, Priv.-Doz., then U. Munich, Priv.-Doz.
1930-32, Rockefeller Found. Fel.: U. Cambridge and U. Rome.
1932-33, U. Tübingen, Priv.-Doz.
1933-34, U. Manchester, Lect.
1934-35, U. Bristol, Fel.
1935-to date, U. Cornell: 1935-37, Asst. Prof.; 1937-to date, Prof.
Interviews
One session, 22 pp. transcript.
Letters to
Møller: 25 Mar 32, 30 Apr 32, (59,4).
Sommerfeld: 5 Jan 33, 15 Jan 33, (29,9).
Letters from
See: Møller, Sommerfeld.
Unpublished material held elsewhere
Bohr Collection, Copenhagen, Denmark. Letters to Bohr, 27 Jul 33, [Jul 33], 19 Aug 33, 30 May 34, 24 Aug 34, 28 Sep 34, 5 letters 1936-54.
Pauli Collection, Zurich, Switzerland. See: Pauli.
Cornell University Archives, Ithaca, N.Y. Personal papers.
Duke University Archives, Durham, N.C. Letter to London, 8 Jun 34. See: London.

BIRGE, R. T. *b.* 1887, Brooklyn, N.Y., U.S.A.
1905-13, U. Wisconsin.
1913-18, U. Syracuse: 1913-15, Instr.; 1915-18, Asst. Prof.
1914, U. Wisconsin, Ph.D. with C. E. Mendenhall.
1918-54, U. California, Berkeley: 1918-20, Instr.; 1920-22, Asst. Prof.; 1922-26, Assoc. Prof.; 1926-54, Prof.
Letters to
Bohr: 10 Nov 19, 28 Apr 22, (BSC 1, 2).
Condon: 9 Nov 26 (50, 7).
Kemble: 6 Feb 23, 22 Mar 23, 4 Jun 23, (50,1); 10 Jul 24, 15 Aug 24, 12 Sep 24, 21 Dec 24, (50,2); 24 Jan 25, 13 Feb 25, 17 Feb 25, 28 Mar 25, 31 Mar [25], 23 Apr 25, 6 Jun 25, (50,3); 8 Jun 25, 20 Jun [25], 22 Jun [25], 25 Jun [25], 6 Jul 25, 12 Aug 25, 21 Aug 25, 23 Aug 25, 29 Aug 25, 5 Sep 25, (50, 4); 20 Oct 25, 29 Oct 25, 7 Nov 25, 8 Nov 25, 17 Nov 25, 30 Nov 25, 21 Dec 25, (50,5); 11 Jan 26, 20 Feb 26, 4 Jun 26, 7 Jun 26, (50,6); 26 Jun 26, 11 Aug 26, 13 Aug 26, 26 Aug 26, 5 Oct 26, 21 Oct 26, 28 Oct 26, 9 Nov 26, 15 Nov 26, 3 Dec

Figure 1.3. Catalogue of the Archives for the History of Quantum Physics (Kuhn et al., 1967).

Yet, apart from the use of the spoken word, the project also thematizes the idea of scientific residue being recognized as a source for the future history of science. As such, it calls for the care and attention of the historian. Within the project, *collecting*, *systematizing* and *situating* the residues within the established archive appeared as a task and responsibility of the history of science.

The resulting "historians' archive" is very different from the "archivists' archive". In contrast to the traditional archive, which guards the "natural" state of affairs, the AHQP represents an *assemblage* constructed and modeled by historians for the historians of the future. As noted in the report, the archive does not adhere to the traditional archival principles of arrangement: "within a single collection, … material was freely arranged for convenient retrieval except in the very rare cases when the original order might itself have historical meaning" (Kuhn et al. 1967, 7). The media strategy of the project is also notable: rather than acquiring (physical) custody over the materials, Kuhn and his colleagues made inventories and microfilm copies. The Archives for the History of Quantum Physics does not guard authentic originals, but brings together copies. The project explicitly disregards the authenticity of the materials and their original order in favor of their heuristic value. It thus alters the very notion of an archive: an archive here is what comes into being when scattered materials (or even their copies!) are brought together to serve as a source of knowledge about the production of science. The archive takes shape as a form of historicization of knowledge, as the fruit of a deliberate effort by historians to accumulate and assemble materials that are deemed important as the primary sources for the history of science.

Following in the footsteps of Kuhn's project, a host of similar initiatives for documenting science emerged and a whole array of scientific archives came into existence. As early as 1968, a similar project was launched documenting the history of computing from the 1930s to the 1960s. By the early 1990s, more than two dozen documentation projects had been undertaken, covering a wide range of topics from the history of Silicon Valley to the history of neurophysiology, and from the history of DNA to the history of the Space Telescope[56].

1.3.2. Archiving the everyday of science

Kuhn's project assembled a collection of the materials to represent the "major episode" in the history of one particular discipline. However, a couple of decades later, as discussions continued and new projects unfolded, there emerged a need for more general guidelines for the archive of science. To address this need, the Joint Committee on Archives of Science and Technology (JCAST) was established in 1979, marking a collaborative endeavor between historians of science and archivists. The committee included representatives from the History of Science Society, the Society of American Archivists, the Society for the History of Technology, and the Association of Records Managers and Administrators.

Operating from 1979 to 1983, the committee set out to achieve three primary goals: (1) to ascertain the extent of available records in the field, (2) to formulate collaborative guidelines for the appraisal and description of these records, and (3) to deepen the understanding of the users and applications of science and technology records. The culmination of JCAST's endeavors was a detailed report (Elliott 1983) offering recommendations on the appraisal and description for the archives of science.

[56] Many of the collections were subsequently digitized and turned into the "digital archives of scientific residues".

As opposed to the thematic project on the history of quantum physics, JCAST aimed to offer general and universal guidelines for scientific archives beyond individual disciplines or time periods. As such, it offered a much broader definition of what counted as science and technology records. JCAST encompassed in its definition all aspects pertaining to research and development, technology transfer, and professional education. However, it deliberately excluded from the consideration medical practices and records related to the use of "widely available technologies." (Elliott 1983, 7).

A key novelty in the JCAST recommendations[57] is the shift in focus towards the daily operations of science. In terms of archival appraisal, JCAST suggests prioritizing the archiving of traces of informal communication and the everyday routines of science over published materials or administrative records. Archiving science is seen here as a documentation of science's everyday life and material practices, rather than a documentation of scientific ideas or bureaucracies.

The report offers a broad and detailed overview of the residues of science to be archived. It recommends preserving, among other things, instrumentation and apparatuses, techniques and their documentation, laboratory notebooks, logbooks, unpublished papers, and other elements that form the "texture of daily activity" (ibid., 27). The report provides individual explanations for each of these types of residues, outlining their potential epistemological value. So, for instance, it delves into the role of logbooks, explaining their function across various disciplines, their connection to research data and scientific instruments.

The report's recommendations rest on the premise that the field of science and technology is so specific that it needs to be explained and elucidated for scholars in humanities and archivists lacking "'a sense for' the nature of scientific and technological research" (ibid., 25). The archive of science should proceed from the nature of the scientific process, and this nature should be elucidated by none other than historians and sociologists of science. Hence, it's no coincidence that the report includes several calls to action specifically targeting historians. It explicitly appeals to historians to take a more active role in the processes of archiving, including appraisal, description, and preservation of scientific materials.

The perspective of the historian of science is contrasted not only with that of the humanities and traditional archival practices, but also with the ideas of the scientists. As discussed in the report,

57 As perceived in reviews of the report (see, for example, Rossiter 1984).

> Many scientists maintain that the published scientific literature normally is sufficient to document research and development. JCAST believes that archivists today should collect fuller record, even though until extremely recently such historical trends as social history, history from the bottom, and quantitative history have played little role in interpreting science. The fuller record will preserve documentary evidence from more that the official perspective; it will incorporate a sociological view, capturing the routine and the revolutionary, the everyday and the profound (ibid., 25).

This is one of the key passages articulating the need to turn to the everyday life of science. What is of interest in this instance is not only the authority of the historian, who is given the main priority in questions about archiving, but also that the report explicitly links the process of archiving with the current paradigms of historical scholarship. It is on the basis of the then turn towards social, quantitative and micro-history that the procedures of archival appraisal need to be modified. It is the new interest of history in a certain kind of sources and the new perspective of the historian that guides the selection of documents as well as their description and contextualization within the archive.

The very idea that the archive could incorporate a "sociological" or any other view is scandalous and inconceivable within the paradigm of the traditional archive. JCAST, though, quite explicitly suggests that the archive should be assembled according to a particular point of view within historical science. The title of the report – "Understanding Progress as Process" – explicitly offers a certain perspective about science and the lens through which it is to be studied. Created by and for historians, the archive no longer claims to provide an "objective" reflection of the past. Instead, it offers a perspective on the past through the lens of the prevailing historical paradigm.

In this sense, the very concept of the archive of scientific residues, as formulated in this work, belongs to a particular paradigm in the history of science. For the archive of residues to "emerge" (or even to be imagined), the history of science first had to recognize the role of everyday (material) objects in the production of knowledge[58]. This shift took place in the 1970s and 80s, a time when the material practices of science captured the attention of historians, and had a rapid, almost instantaneous effect on the conception of scientific archives[59].

---

[58] For some recent historical studies focusing on everyday material objects and practices in science see Waquet 2015; Werrett 2019. On the material practices of knowledge production in humanities, see Lamy, Bert 2021.

[59] Concurrent with the growing historical interest in the scientific residues, the 1980s witnessed what is termed the "patrimonial turn" in science. While in history, residues began to be viewed as objects of knowledge, the patrimonial turn approached them more as objects of heritage. This shift towards a patrimonial perspective in science is particularly emphasized in the French tradition (Boudia et al. 2010, Boudia, Soubiran 2013).

1.3.3. Computational thinking and archive modelling

The third episode I would like to touch on is how the historical/ archival perspective changes with the appearance of the computational history of science. Digital forms of "gathering, preserving, and presenting the past," quoting the title of the renowned textbook by Daniel Cohen and Roy Rosenzweig (2005), lead to yet another shift in the contours of the archive as it is articulated and assembled by historians.

While in the previous two cases I examined specific archiving initiatives, in this instance, I will focus on the theoretical outlines of the archive as they are formulated in digital history (of science)[60]. This discussion is continued and animated with many examples in the subsequent chapters.

Computational historical research operates not with sources, but with data (and preferably "big" ones). Once converted[61] into data, sources, firstly, can be manipulated and processed by machines, which opens up opportunities for various types of "distant reading" and other forms of data-driven research. Secondly, unlike sources, data are (or at least should be) open, sharable, and interoperable. Therefore, data from a particular research project should no longer be relegated to obscurity. Instead, they can be brought into the spotlight, re-interpreted and re-used, published, exhibited and framed as archives.

If, in the wake of the digital turn, still not "everyone is an archivist" (Ketelaar 2006), every digital historian certainly is. The scientific archive is being increasingly recognized as a *data archive* for the digital history of science. Such archives are meant to aggregate "different types of data, revealing networks of connections and enabling comparison across far more types of data than would otherwise be possible" (Laubichler, Maienschein, Renn 2013, 122-123).

This emphasis on data accumulation, aimed at overcoming the fragmentation of sources, was partly foreseen already in Kuhn's early project. However, in the projects of digital history of science, it is meant to take on a dramatically different, *macroscopic*, scale. Big data

---

[60] The project of computational history of science was articulated in two seminal articles by Manfred Laubichler, Jane Maienschein, and Jürgen Renn (2013 and 2019).

[61] The transformation of sources into data, a process known as 'datafication,' involves a series of technical procedures – digitization, conversion into standardized digital formats, normalization, annotation, and the addition of metadata. That said, datafication is far from neutral or natural. For a more detailed discussion of the data life cycle in the context of a history of science project, see Damerow and Wintergrün 2019.

equips the historian with a new toolkit, a "Macroscope" (Graham et al. 2015), which enables one to "grasp the incredibly large".

The "macroscopic" gaze unveils a new phenomenon to be explored through the aggregation of data. As Franco Moretti, outlining his project of uncovering "the great unread" through distant reading, notes:

> [I realized] what a minimal fraction of the literary field we all work on: a canon of two hundred novels, for instance, sounds very large for nineteenth-century Britain…, but is still less than one per cent of the novels that were actually published …a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn't a sum of individual cases: it's a collective system, that should be grasped as such, as a whole (Moretti, Piazza 2007, 3-4).

The digital archive is meant not simply to bring together a multitude of disparate sources, but to document such a "collective system", *a whole* not reducible to the sum of individual testimonies (or residues). So, for instance the "Expeditions and Discoveries" digital collection[62], which gathers residues of scientific expeditions between 1626 and 1953, aims to capture the expedition practices at large, going beyond the scope of any specific one. Similarly, the "Sound and Science: Digital Histories" collection[63] seeks to reflect the overall dynamics of intersections of science and sound.

It is these new economies of scale that allow for bringing in new research questions. So, according to Manfred Laubichler, Jane Maienschein and Jürgen Renn, computational history of science enables performing "*longue durée* assessments of the evolution of knowledge on a global scale, detailed analysis of the nature of scientific transformations and revolutions, and comparisons across wide temporal and spatial scales" (2013, 120). "Macroscopic thinking" places emphasis beyond individual items, on the patterns and connections they collectively form. The issue at stake is hence not only the accumulation of knowledge, but also its fundamentally new structure, in which the relationships and links between records become more important than the archival items themselves.

Beyond aggregation, the digital history (of science) brings in the idea of data *modelling* within the archive. For data (rather than things) to speak, they need to be structured according to certain knowledge models or information systems: indexed, classified, databased, modelled (in complete accordance with the information paradigm discussed in chapter 1.1).

---

62 Harvard University Library, [https://curiosity.lib.harvard.edu/expeditions-and-discoveries]

63 [https://soundandscience.net]. A collaborative project between several institutions, initiated by Max Planck Institute for the History of Science.

Accumulating numerous pieces of data, the archive is to situate and arrange them in a certain way, to structure and model them, to outline the possible types of links between them.[64]

Within the Digital Humanities field, modelling appears as both a method and a focus of inquiry. As a method, it encourages historians to engage with databases and knowledge systems, to construct ontologies, develop metadata standards, and investigate the possibilities of Linked Data for historical datasets (Meroño-Peñuela et al. 2012).

But more than that, modelling – creating a model or a (simplified) representation of a phenomenon – appears not just as a technical procedure, but as a conceptual part of research, as "thinking-in-doing"[65]. When modelling data, the historian thereby proposes a certain representation and *interpretation* of the phenomenon, encoded into technical systems and algorithms. Thus, shaping and situating data within an archive appears as a conceptual task, which sets the horizon for their interpretation.

Modelling thus stands as a new form of (conceptual) intervention of the historian into the archive. In the traditional archive, documents are stored according to its inherent logic. Here historians extract some materials in order to then (re)situate and (re)contextualize them within their historical research. The digital history archive introduces a different epistemological framework. In this case, historians themselves model the archive, conceptualize and annotate its structure, outline the network of relationships and connections between items, so that the archive can subsequently produce (historical) knowledge. Such an archive is no longer just an inventory preceding research. Instead, it becomes an integral part of the research process or even one of its outcomes, on par with publications. Canonical early examples of such projects in the field of Digital Humanities include the Walt Whitman Archive, the William Blake Archive, and the Rossetti Archive[66].

Referring to such digital projects as "archives" is not without controversy. I have previously mentioned some objections from archivists in this regard, particularly those from Kate Theimer (2012), who argues that such collections cannot be called archives since they do not adhere to traditional archival principles. Within the Digital Humanities field, there is an ongoing discussion, revolving around the different terms to denote such forms of historical inquiry: *digital archives*, *databases*, *thematic research collections*, *digital critical editions*. The

[64] For various strategies and forms of data modelling, see Flanders, Jannidis 2015.
[65] Willard McCarty (2005) formulates an entire philosophy of modelling, positing it at the very core of the theoretical reflection in computational humanities.
[66] In the history of science, many examples of such initiatives are collected on the Digital History and Philosophy of Science Consortium website [https://digitalhps.org/projects]. Some of these are discussed in later chapters.

boundary between the genres is often quite subtle: the same digital project can be described through several categories at once. Kenneth Price examines these genres through his own project, the *Walt Whitman Archive*, demonstrating how it sits in between these categories, not fully fitting into any one of them, yet finding points of contact (or tension) with each (Price 2009).

As opposed to the archive, the database (at least in the strict sense) is rather a technical term referring to collections of structured data that are controlled by specific management systems[67]. The critical edition, in contrast to the archive, generally pertain to *texts* (rather than material objects), enriched with academic and analytical commentary (McGann 2010; Pierazzo 2016). Perhaps more attention should be paid to the relations between digital archives and thematic research collections. *Digital thematic research collection* refers to digital aggregations of primary sources gathered around a particular topic and intended for research (Palmer 2004). The relationship between the two terms remains somewhat ambiguous: is a research collection a sub-type of an archive, or are these two intersecting genres? In the field of Digital History, this is purely question of naming and no consensus is reached (Gerber 2023). For archivists, on the contrary, the distinction is essential, since it reenacts the traditional divide between organic and artificial collections. (In the same logic, any collection assembled by historians – for instance, the Archives for the History of Quantum Physics – would not be considered a proper archive). However, if, as discussed in chapter 1.1, the distinct characteristic of the digital archive is its inability to perform traditional archival principles, such as provenance or original order, then such a radical demarcation becomes less meaningful. Therefore, throughout this work, I consider thematic research collections as a subtype of digital archives.

If, as Michel de Certeau (1992 [1975]) wrote, history is an interpretative practice, then the archive, formed under its supervision, is one of its modes of interpretation. Taken together, the three episodes examined in this chapter demonstrate how the current state of historical science influences the formation of archives. From Kuhn's project to contemporary digital endeavors, archives are assembled, appraised, and described in accordance with specific conceptions of knowledge. Subsequently, these archives will, in their turn, shape the ways in which knowledge about the past will be produced and understood in the future.

---

67 The metaphorical meaning of the database will be to some extent discussed in the chapter 2.1.

1.4. Collection of collections

1.4.1. Digital archive of scientific residues: a definition

Tracing how the concept and contours of the digital archive crystallize, and how the notion of an archive of scientific residues takes shape, this chapter finally comes close to the phenomenon at the center of this study – the digital archive of scientific residues. This phenomenon resides at the convergence of the two shifts. The first shift – from the traditional to the digital archive – is primarily a technological transition, driven by the emergence of new infrastructure, media, and digital practices. In this regard, the most straightforward way to define a digital archive is through its media and technologies. Digital archives are essentially digital aggregations of materials, complemented by information infrastructure components, such as indexing or metadata systems.

Conversely, the turn to the residues of science is rather conceptual one. The idea of the scientific archive as an archive of residues takes shape in the wake of burgeoning interest in the everyday practices of science within historical research. As we saw in the previous chapter, the history of science is gradually turning towards the concept of scientific residue as a source to be collected, preserved, and situated in a certain way. The archive of residues can be conceptualized at the point when history and sociology of science begin to inquire into the workings of "science in action" (Latour, 1988), exploring how knowledge is produced through specific material practices.

The identification (or *construction*) of the digital archive of scientific residues as a research object is thus grounded in two basic distinctions: firstly, between the traditional and the digital archives, and secondly, between the scientific library, which stores publications and data, and the scientific archive, which preserves scientific residues. Synthesizing these two lines into a more formal and broad definition, *the digital archive of scientific residues* represents digital aggregations of the material remnants of knowledge production.

This concise formulation also calls for a number of clarifications. In terms of technology, these "digital aggregations" should be equipped with certain infrastructures adopted in the field of digital preservation: metadata systems, archival descriptions, digital formats, knowledge organizations systems. From a historical science viewpoint, these "material remnants" ought to be considered by the history of science as (potential) primary sources. That said, the digital archive of scientific residues, as defined here, encompasses both digitized and born-digital archives, both institutional and thematic research collections, and those created by both professional archivists and community archives.

This (quite broad) definition is further operationalized through assembling *this dissertation's corpus of archives.* Much like a text corpus, this corpus is a curated collection of selected archives, designed to exemplify the phenomenon under study. Establishing a corpus of archives in this instance is intended to map out the broader phenomenon of the digital archive of scientific residues, transcending isolated examples. This approach allows for a comparative analysis of individual collections, not in isolation or as absolute entities (which could lead to somewhat speculative or impressionistic conclusions), but in relation to and against a broader backdrop. It thus allows either challenging or reinforcing hypotheses and conclusions drawn from specific cases. Having a corpus of archives one can also perform its *distant reading* (Moretti 2013) discerning some common trends and patterns in the way in which the digital archives manifest themselves.

The analysis of this corpus is the focus of the upcoming part, while the remainder of this chapter is dedicated to a formal description of the corpus, as required by corpus linguistics.

1.4.2. Collecting the corpus

Target population

The corpus is meant to represent the digital archives of scientific residues as conceptualized above. The target population *does not* include "institutional research repositories", aggregating what is considered to be the research output of the institutions, i.e. publications and data. That said, I incorporated into the corpus institutional and museum archives and "research" collections created within the scope of computational history of science projects.

In terms of thematic scope, I excluded industry and business archives, medical collections, and archives representing "campus life" from consideration. Instead, my focus was on collections related to "research and development", largely aligning with the definitions provided in the JCAST guidelines.

Finally, my examination was limited exclusively to open access digital repositories.

Searching for the archives

In compiling the list of archives[68], I navigated a variety of sources: digital library guides, project indexes in the history of science, general lists of scientific research institutions, and archival search engines.

[68] The list of archives is available in Appendix 1 (Table 1). All other types of data and statistics that I will refer to in the course of the analysis are available at [https://zenodo.org/records/10548106].

In the absence of comprehensive indexing of digital (archival) collections, the search was carried out manually, inevitably incorporating numerous random factors. Its scope was, in one way or another, limited by my location and my language proficiency (which are also factored into search algorithms), leading to an inherent bias towards North American and European sources[69]. For instance, I was unable to find any Chinese collections of scientific residues, though it is plausible that such collections exist.

That said, manual searching had its advantages. Over its duration (which lasted nearly two years), I was able to review a wide variety of archives, to observe both the 'norm' and some deviations from it, and to include examples of both in the corpus.

The unit of analysis

Archival repositories vary greatly in terms of their scale and diversity. Some are relatively small and homogeneous, while others amalgamate many and diverse collections. Take, for example, the Cambridge Digital Library, which comprises nearly a hundred collections, ranging from Medieval medical recipes to caricatures, and from Darwin manuscripts to Southern African collections.

To enable comparison across archives and repositories, I have selected *the collection* as the basic unit of analysis for the corpus. For example, the corpus includes two collections from the archives of the University of Cambridge: those of the Cavendish Laboratory and of the Royal Greenwich Observatory. It is important to take into account, however, that collections can be organized on different grounds: a collection in the traditional archival sense (a set of artifacts ingested from a single source, e.g., the "Frederick Winslow Taylor Collection" (Stevens Institute of Technology)); a thematic collection curated for a digital archive (e.g., video collection of the American Philosophical Society); or a collection organized according to a museum's exhibition layout, such as "Mathematics: the Winton Gallery" (Science Museum Group).

Sampling procedures and the balance of the corpus

In the field of corpus linguistics, a corpus is usually established by defining a sampling frame. The sampling frame represents a comprehensive list of all potential sampling units from which a corpus is systematically constructed, often through a process of random selection. An

[69] That said, this "bias" also reflects the observed disproportion in the digitization of materials between the Global North and the Global South (cf. Zaagsma 2023).

example of such a frame could be a catalog of all books published within a specific time frame in a particular country.

Since the practices of indexing digital collections and projects are still not on par with bibliographic standards, in my case, outlining a comprehensive operational sampling frame was impossible. Lacking a complete list of all available collections, it is difficult to assess if the selected sample represents the whole. This issue poses the most significant challenge concerning the representativeness of the corpus.

Therefore, my strategy was to make the corpus as varied as possible. In compiling it, I deliberately included both well-recognized archives known for their exemplary practices, like those from the Library of Congress or Bibliothèque nationale de France, and smaller, more localized, and somewhat "exotic" collections, such as *it-museum*, a private Russian collection focused on the history of Soviet computing, or *Bitsavers.org*, a collaborative archive of software. This approach was taken to observe not just what might be considered the conventional "norm" but also to explore the outliers and deviations from it.

In a more stringent sense, the degree to which a corpus reflects the variability within a population is typically characterized as the balance of the sample (Biber 1994, 243). This aspect forms the second defining criterion of its representativeness. A corpus is considered balanced if it comprises a diverse range of categories of 'texts'. Consequently, establishing a corpus requires an initial classification of the target population to determine the categories that should be included. In what follows, I will detail these categories.

Primarily, in terms of institutional framing, the corpus encompasses both digital collections of scientific institutions and museum collections focusing on the residues of science, as well as archives established as part of projects in the computational history of science. Notably, the archives of scientific institutions constitute just over half of the sample (Fig. 1.4.).

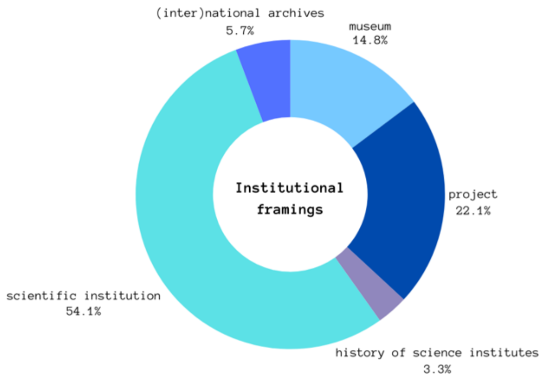


Figure 1.4. The percentage distribution of collections in accordance with institutional frameworks.

Additionally, the corpus is broadly distributed in terms of geography (Fig. 1.5). It includes archives from a variety of countries, with the majority being from the United States (constituting just under half of the corpus), followed by Europe (primarily Germany and France), and the United Kingdom. It also incorporates six Canadian archives and one collection each from Brazil, Russia, and Australia. As mentioned earlier, this quite narrow geography is attributable, on one hand, to the mechanisms of search and, on the other, to the general disproportionality in the policies governing the digitization of archives.

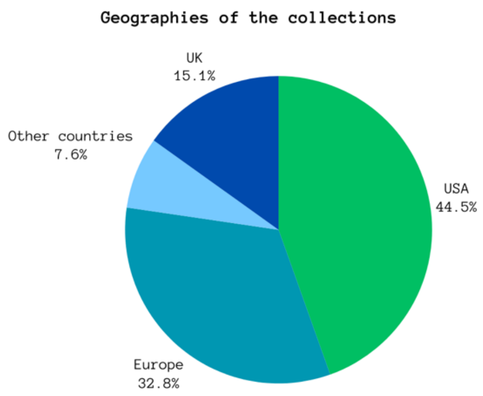


Figure 1.5. The percentage distribution of collections in accordance with their geographies.

Further, the corpus encompasses a variety of collection types (Fig. 1.6), including personal collections, thematic (subject-based) collections, collections of scientific instruments and technologies, as well as institutional collections representing the history of an institution. These latter, in turn, represent the histories of a wide array of institutions including laboratories, universities, observatories, international associations, scientific societies, and research organizations.

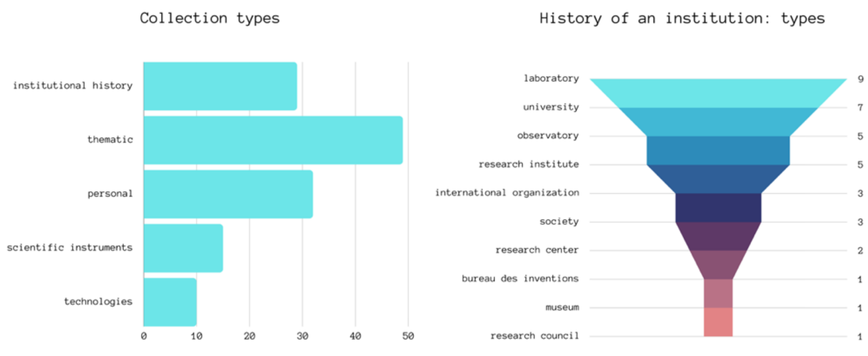


Figure 1.6. On the left, the distribution of collections by type is shown (the x-axis displays the number of collections in absolute figures). On the right, the distributions of institutional history collections by type of institution.

The corpus also spans a diverse range of disciplines, from physics and life sciences to social sciences and scientific management (a complete list of disciplines and their distribution is presented in Fig. 1.7). Notably, a significant portion of the collections represents multiple disciplines simultaneously.

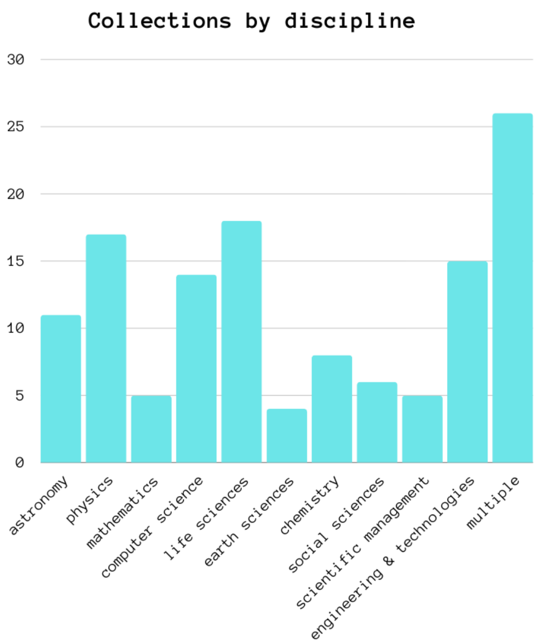


Figure 1.7.The distribution of collections by discipline

Next, the corpus brings together collections representing different genres of items (Fig. 1.8): texts (laboratory notebooks, drafts, correspondence, notes, papers), images (scientific visualizations, photographs, technical drawings), objects (scientific instruments, machines, equipment), sound (audio recordings of presentations and workshops), video (interactive visualizations, video recordings, clips) and web materials (software, emails, websites, blogs). Since most collections exhibit multiple genres, in this case, I made sure that all *genre types* (not collection types) were represented in the corpus.

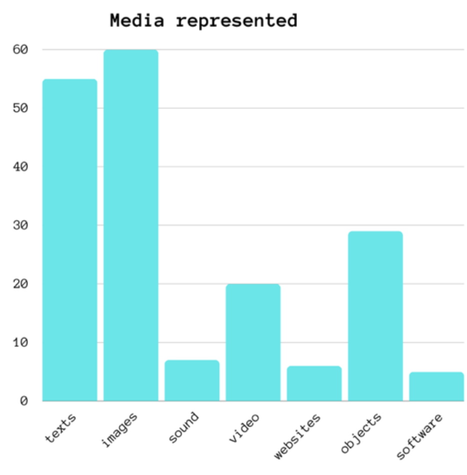


Figure 1.8. The distribution of genres of items by collection

Finally, the corpus covers collections featuring digitized as well as born-digital materials (Fig. 1.9). A significant proportion of them, however, contain both.

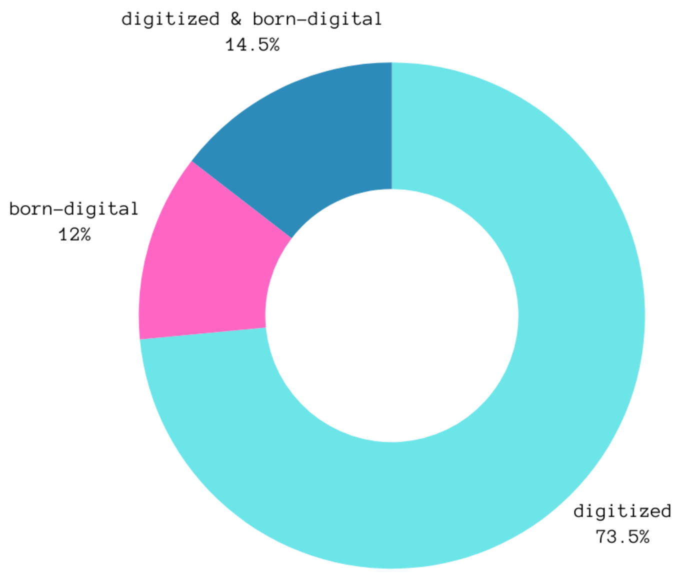


Figure 1.9. The percentage distribution of digitized and born-digital collections.

The resulting "collection of collections" encompasses one hundred eighteen archives across countries, institutions, formats and scientific disciplines.

The annotation and processing of the corpus were carried out manually using several Excel spreadsheets. The statistics were for the most part calculated with the help of Excel statistical tools. Whenever coding was required, I described the procedure in the relevant section. For all other instances, I referred to the data from the specific Excel document, providing a reference link.

The next part proceeds with an analysis of the corpus, focusing on how digital archives represent and situate scientific residues. Meanwhile, the remainder of this chapter will set out some preliminary observations regarding the corpus.

1.4.3. Digital selectivity

In engaging with a corpus, it is important to understand that digital collections maintain intricate and non-symmetric relationships with their physical collection counterparts. What gets digitized often only accounts for a small segment of the entire archival record. For instance, the digital collection of Alexander Graham Bell at the Library of Congress[70]

70 [https://www.loc.gov/collections/alexander-graham-bell-papers/]

showcases 4,695 items from a physical collection of 145,000 items. Out of 100,000 documents representing the inventions of the General Electric Research Laboratory[71], only 1,758 have been digitized and published online. Even the digital collections of Albert Einstein, Alan Turing, and the extensive Thomas Edison collection – the result of decades of work – still only display a portion of the documentation available in the physical archives.

In addition to the traditional archival appraisal, the digital archive presents archivists with yet another choice: what is to be transferred from the physical archive to the digital one. The selection process thus becomes twofold: what gets archived and what gets digitized[72] (and therefore exhibited and *seen*).

This "digital selectivity" is influenced by many, often pragmatic, factors. For instance, funding plays a significant role. Digitization, still a relatively costly endeavor, is often financed by private foundations and grants. For example, the digitization of a relatively modest collection, *the Papers of Georg and Max Bredig* (Science History Institute holdings[73]), which contains around 3,000 documents, required a grant amounting to almost $200,000[74]. The preferences and focuses of financial stakeholders (e.g., grant-giving organizations) directly impact the decisions regarding which collections are chosen for digitization and what segments of these collections are prioritized (Ogilvie 2016).

Further, digital selectivity is deeply connected with wider political dynamics. Gerben Zaagsma (2023), in outlining the politics of digitization, points to various types of inequalities: the disproportion between digital archives in the Global North and Global South, postcolonial legacies, and issues of political censorship. These factors of digital disparity often result in archival silences. In an era of pervasive digitization, what fails to be included in the digital archive becomes even more invisible than it was before the digital turn. A separate issue is that many collections lack information on which specific part of the physical collection was digitized, let alone the rationale behind these decisions.

In addition, there is also a broader cultural and historical perspective to consider: public discourses on the past and politics of memory. For instance, the digitized Freud collection at the Library of Congress[75], represents only the artifacts directly associated with Freud,

---

71 [https://nyheritage.org/collections/ge-research-lab-photographs]

72 In the case of born-digital collections, the selectivity is usually integrated into automatic archiving systems, such as Internet archive crawlers. For the special case of "total archive" preservation, see below Part 4.

73 [https://digital.sciencehistory.org/collections/qfih5hl]

74 [https://www.sciencehistory.org/about/press/bredig-clir-grant/]

75 [https://www.loc.gov/collections/sigmund-freud-papers/]

excluding those related to his family. An even more intriguing case is the Cavendish Laboratory collection[76]. When choosing photographs to digitize, the curators of the collection took cues from the images that had been featured in previously published accounts of the laboratory's history. In other words, the digital archive, in its selection process, was steered by an already existing historical narrative and replicated its logic.

Nonetheless, the rationale behind digital selectivity is not applicable in collections that inherently lack physical equivalents. This is particularly true for born-digital collections and most of the archives established through computational projects.

1.4.4. Socio-technical interplay

The digital archive is a socio-technical ensemble, the result of the interplay of people, technical infrastructures, standards, protocols and institutional policies.

On the technological front, it is shaped by the synergy of various interrelated, multi-layered and distributed technologies (Blanchette 2011). This includes workflow systems, digital repository systems (such as *Zenodo* or *Islandora*), content management systems (*ContentDM*, *Omeka*, *Drupal*, *DSpace*), digital asset management systems (*Fedora*), interfaces, digital exhibition platforms (*Omeka*). These systems are typically spread across different locations, managed by different entities, and operate interdependently to support the complex functionalities of digital collections.

All the examples listed come with their own affordances and constraints: various repository and content management systems support different metadata formats (such as METS or Dublin Core) and protocols (e.g. OAI–PMH or IIIF), have different systems for creating item identifiers, and offer different features for exhibiting archival items. In addition, some repositories operate on unique, customized systems, developed through collaborative efforts and negotiations among archivists, programmers, designers, and data managers. Moreover, the archives implement different protocols, metadata standards, digital formats, controlled vocabularies and indexing systems.

Added to this intricate and layered technical infrastructure are the dynamic social and socio-technical interactions among professionals engaged at different stages of managing the collection. This group spans a wide range, from archivists, digital curators, and information managers to designers and back-end developers. Not to mention the differing institutional

[76] [https://cudl.lib.cam.ac.uk/collections/cavendish/1]

archival policies, financial considerations, and many other factors influencing the form a particular collection will take.

As a result of this entire orchestration, a digital archive comes into being as a unified system with which the user interacts. The way the collection is organized, the way it situates the artifacts and presents them to the user is the fruit of these numerous interactions.

One way to analyze the archives is to delve into the unique dynamic behind a particular collection and meticulously trace how, through it, the collection takes shape. In this case, any given collection will appear as one of a kind, driven by a unique combination of factors and interactions. Another approach is to analyze a digital collection as a finished result, as it is presented to the user. While sacrificing the variety of factors and contingencies "behind the scenes", this approach allows comparing collections with each other and identifying tendencies. It is this method that I have chosen to adopt given the objectives of this study. Inquiring about how the digital archive represents scientific residues, my focus is not on the reasons behind each specific individual representation. Rather, I am interested in the general and typical aspects, those that distinguish the digital archive of science as a distinct genre, medium, or *dispositif*.

Therefore, in analyzing the collections, I do not deal with the individual actors behind it: I question neither the decisions of archivists, nor the policies of an institution, nor the affordances of any particular technology. In this sense, one could say that distant reading analysis sacrifices what Yanni Loukissas (2019) called the locality of data – the specific actors and choices behind particular datasets and digital representations. While this approach does not afford to examine the specificity of each particular collection, it offers another form of interpretative examination: it seeks to identify certain *regularities* in how digital scientific archives manifest themselves across specific systems and institutions[77].

1.4.5. Sustainability and obsolescence

The final observation concerns the particular dynamics of sustainability and obsolescence inherent in the digital collections.

In physical archives, primary attention is given to the inevitable aging of archival objects, whereas the obsolescence of the *archive* (its buildings, metadata systems, paper cards, etc.) is considered a secondary concern. In contrast, in the digital archive, it is not the archival

77 While Part 2 does not take into account the locality of the data, Part 4, focused on the analysis of one specific archive, on the contrary, is particularly attentive to its social and cultural settings .

objects themselves that are vulnerable to aging and change, but rather the entire archival layered infrastructure: interfaces, content management systems, hosting platforms and so forth.

The infrastructure of a digital archive lacks the stability of its physical counterpart. Digital archives require constant updating and care of their numerous software layers. This ongoing need to keep up with technological changes often results in completed digital projects being "frozen", abandoned, and ultimately turning into "digital wastelands" (Barats, Schaffer, Fickers 2020).

Notably, the life cycle of some collections in the corpus turned out to be shorter than the time needed to prepare this dissertation. In the time span from assembling the corpus to writing this text, some collections ceased to exist (such as Meteorological Observations collection of the Royal Society[78] (Fig. 1.10)), others underwent complete transformations (Caltech Images collection[79]), changed their indexing and metadata systems (Lick Observatory Photographical Archive) or moved to another URL ("Sound and Science" digital project[80]).

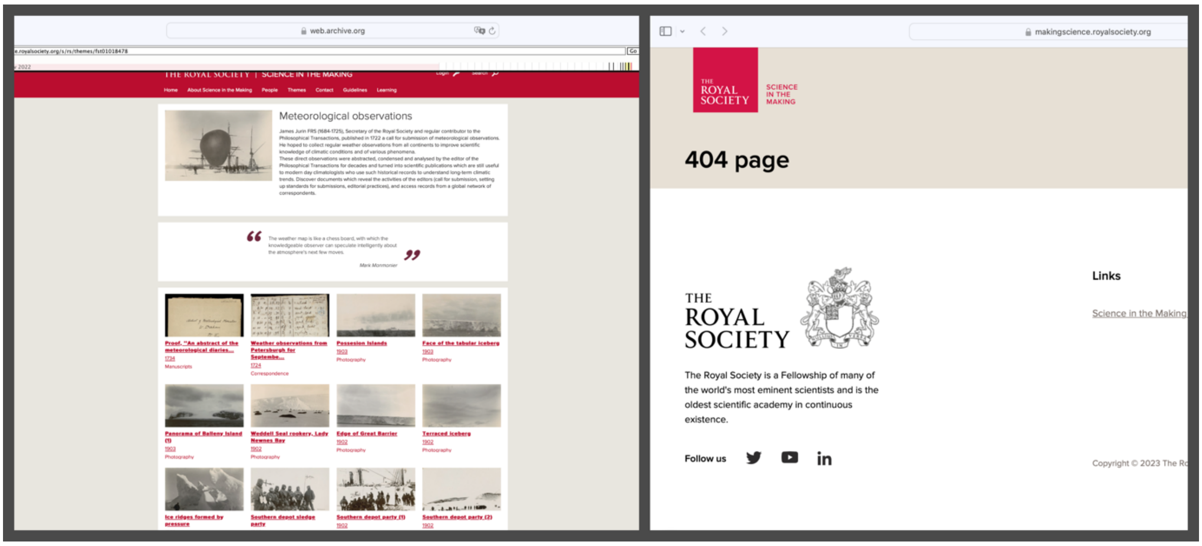


Figure 1.10. The Meteorological Observations collection in 2022 (left) and the current (November 2023) collection page.

This dynamics of change and obsolescence carries one significant implication for the analysis of the corpus. Namely, it deals with an ever-changing and unstable object, one that constantly alters its contours. Therefore, my analysis captures the actual state of the collections as of 2022-2023. It is entirely possible that by the time of the dissertation defense, it may be describing an entity that no longer exists in its current form.

78 [https://makingscience.royalsociety.org/s/rs/themes/fst01018478]
79 [https://collections.archives.caltech.edu/repositories/2/resources/219]
80 Its old URL: https://soundandscience.de/browse-objects. New URL: https://soundandscience.net

Conclusion

The transition from the traditional to the digital archives is often articulated as an *ontological* shift: from a stable, integral, and tangible archival object to an ephemeral digital entity, devoid of its (former) materiality. *Le goût de l'archive*, as depicted by Arlette Farge (1989), the historian's intimate and tactile engagement with sources in a physical archive, is no longer in place, being substituted by *le goût du numérique* (Clavert, Muller 2018).

However, what I intend to emphasize by situating the digital archive between the archival and information paradigms is the *epistemological* shift that comes with the digital turn. The digital archive, quite evidently, loses the "naturalness" and organicity inherent to its traditional counterpart, stepping instead into the realm of *knowledge representation*. In it, information technologies, indexing systems and semantic models take center stage, representing and situating archival items in certain ways and implying certain retrieval structures. We observed this transition in the endeavors of computational historians, which turn the archive modelling into a conceptual task, and especially in the pragmatic/hermeneutic reimagining of the archive within the OAIS standard.

As OAIS suggests, the archive henceforth preserves not so much the records themselves as the horizons of their understanding. The plethora of terms introduced by the standard – ranging from "information" to "long-term" and "Designated Community" – reflects a shift towards the de-substantialization of the archival item (which gains meaning only in use or when understood), and of the very idea of preservation (which can only be for someone). Ultimately, this redefines the archive itself, which no longer safeguards content but, instead, is engaged in the continuous *updating of representations based on the current state of affairs*.

In this light, the notion of "representation information" becomes particularly telling. It emerges as the foremost method for contextualizing an archival item, replacing provenance and "original order". What we observe is a shift in the very *arché* of the archive, "the first law of what can be said" (Foucault 1972 [1969], 129). As a form of knowledge representation, the digital archive has its own epistemological affordances and constraints, renders possible certain historical interpretations while precluding others.

Digital archives of scientific residues, in this context, present an especially intriguing case for examination. The material residue of scientific practices was at some point "discovered" and thematized by the history and sociology of science as an object-knowledge, bearing testimony to certain facets of how science operates. If this is the case, then the quantitative landscapes of residues, made accessible, structured, and contextualized through the digital infrastructures, might offer some novel perspective on science-in-the-making.

To frame it as a research question, which drives all the subsequent chapters: how does the digital archive of scientific residues shape "what can be said" about the past of science?

Part II. Archival Representations: how do digital collections bring order?

To explain his much-celebrated maxim “medium is the message,” Marshall McLuhan cited the example of a simple light bulb. The light bulb contains no “content” in the usual sense, yet still conveys a message: it forms our environments, influences our actions, attitudes, schedules, arrangements of everyday life. “Whether the light is being used for brain surgery or night baseball is a matter of indifference”, McLuhan claims, as a medium, however, it “shapes and controls the scale and form of human association and action” (McLuhan 1964, 23-24).

But what happens when this very light bulb, a “pure medium” with no content, makes its way into the (digital) archive? Consider, for instance, these representation of bulbs in the Science Museum Group, the Harvard collection of scientific instruments and Oxford’s History of Science Museum (Fig. 2.0.1):

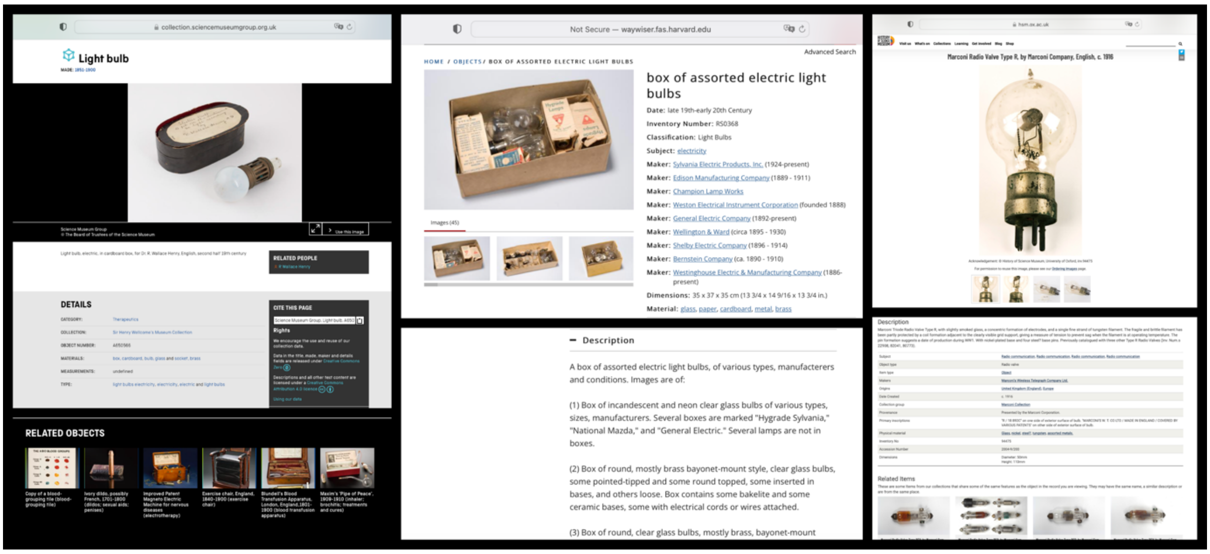


Figure 2.0.1. The representations of light bulbs in three digital collections: the Science Museum Group, the Collection of Historical Scientific Instruments at the University of Harvard, History of Science Museum at the University of Oxford.

In the archive, the light bulb is deprived of its mediality. It neither lights up nor flickers, it no longer creates an environment, but itself becomes the “content” of another medium – the digital archive. The way we make sense of the bulbs now depends entirely on their presentation in and through the archive. The first bulb is *represented* as a medical instrument of the early 19th century; a set of bulbs from the Harvard collection is *described* in terms of their manufacture; the Marconi radio valve is *introduced* as an instance of a certain bulb type. Here what McLuhan called “the matter of indifference” suddenly comes to the fore: what the context of the light bulb is, how it is being described and represented.

In the course of this brief exercise in media theory, it becomes apparent that the order of the archive – the way it represents objects and gives meaning to them – *is the message too*. The archive displays and situates the object, defines its relationships, assigns it to categories, frames it with words, descriptions, numbers, dates, as well as with other artifacts. In so doing, it shapes the background context of the thing, sets the framework for the user's perception and interpretation of it, and defines the ways in which the artifact in question relates to other objects in the archive.

The main purpose of this part is to problematize and question the digital archive as a medium endowing scientific artifacts with meaning. How do the digital collections represent, order and give meaning to the residues of science? How do they mediate our encounter and engagement with the past of science?

In posing the question of mediation, this part in a certain sense follows Library and Information Science (LIS)'s view of the archive as a form of representation and organization of knowledge (Stock 2015; Harmelen et al. 2008). However, if LIS mainly articulates the pragmatics of different forms of knowledge organization – its affordances for information retrieval – I would rather be interested in the way history is represented and imagined through the agency of the archives. My analysis in this sense will be close to a phenomenological one: what do we encounter in the digital archives as users and as historians, looking *at* the collections rather than looking *for* a particular artifact? What does the archive of science look like as an infrastructure? What does it communicate? In what ways does it characterize and situate the residues of science?

In this sense, this part is more in line with those historical studies that question the epistemological implications of different forms of representing history for historical research. This approach to historical poetics is practiced, for instance, by Stephen Bann (1986; 1990) analyzing the structure and forms of display in historical museums, narratives and art exhibits. Regarding digital collections, a similar range of questions is addressed by Joshua Sternfeld, who introduced digital historiography – "a new interdisciplinary theory dedicated to the construction, use, and evaluation of digital historical representations" (Sternfeld 2011, 547). The key question of the new theory as defined by Sternfeld is echoed in this part: How are historical representations constructed and contextualized?

In this analysis, I will rely on the corpus of 118 collections, outlined in the previous part. Analyzing more than a hundred collections, I thereby turn them themselves into material for distant reading, assuming that we can read archives and not only *through* archives. That is, through distant reading, I propose to shift the lens and look not at the objects stored in archives, but at the medium itself: to investigate how archives are organized, how they represent objects, how their classification languages are organized. Quantitative methods will offer a way to capture tendencies to be further interpreted and complemented by specific cases. Taken together, distant and close reading will provide a state-of-the-art view of current digital collections of science.

Outlining this state-of-the-art, I will successively examine different facets of the archival order. I will begin with some general remarks on the structure and order of the digital archive as opposed to that of the traditional archive and the narrative as forms of representation. I will then turn to the question of how the visual *dispositif* of digital collections is arranged looking at the ways in which the collections position and display scientific artifacts. The third chapter will focus on the question of digital connections and the ways in which the digital archive establishes links between the objects of science. Finally, the last chapter will discuss the structures and languages of archival classifications.

2.1. Towards the digital order

Digital collections do not tell stories. This is the point at which all the (numerous) attempts to define the digital archive as opposed to the narrative converge. This is not to say that this point is brand new: already Arlette Farge in her classic analysis of the traditional archive noted that "the archive does not write the pages of history"[81] (Farge 1997 [1989], 13). Nevertheless, the digital archive, no longer anchored in the materiality and physicality of things, raises anew the question of its relation to both history and story. What is at stake is the question of how we encounter the past and make meaning of it through the digital archive. In what way, indeed, does the archive represent the past, if it neither promises the privilege of "touching the real"[82], nor relies on the main form of assigning meaning: the narrative?

It is not surprising, therefore, that the humanities reflection on the digital archive tends to proceed from its confrontation with the narrative. In what follows, I will first deal with the three main arguments behind the opposition of archive and narrative: "structural", "pragmatic" and "technological" ones. Then, using one particular collection as an example, I will present a more nuanced analysis of the way the digital archive situates the historical artifacts as opposed to both the narrative and the traditional archive.

---

[81] My translation from French.

[82] As noted by Farge, "by unfolding the archive, we obtained the privilege of 'touching the real'" (Farge 1997 [1989], 18).

2.1.1. Digital archive vs. narrative: the three arguments

1 Structural argument

In Lev Manovich's famous phrasing, the narrative and the database[83] are "natural enemies" (Manovich 1999, 85), two symbolic forms operating on fundamentally different grounds. In his analysis, Manovich focuses primarily on the structural differences between the two forms. For him, collections stand in opposition to the sequential structure of the narrative: they "don't have beginning or end; in fact, they don't have any development, thematically, formally, or otherwise which would organize their elements into a sequence" (Manovich 1999, 80)[84]. In fact, in the database, all entities are separate and equal in their rights. Whereas the narrative is linear, arranged and produces cause-and-effect connections[85], the database is fractured, fragmented, and "represents the world as a list of items which it refuses to order" (Manovich 1999, 85). While in the narrative nothing is accidental and all is subordinated to its coherence – the development of the plot – the database emphasizes choices, alternatives, various possibilities, the non-linearity of connections between items. This structural distinction, conceptualized by Manovich in the Saussurean terms of syntagm and paradigm, marks a radical break between the database and the narrative as two symbolic or cultural forms[86]. According to Manovich, it is the logic of the database, as a new cultural form, that increasingly structures our experience and worldview, so that the "the world appears to us as an endless and unstructured collection of images, texts, and other data records" (Manovich 1999, 81).

---

[83] Manovich is not operating with a strict (computational) definition of the database, understanding it in a broad sense as "any collection of items on which users can perform various operations: view, navigate, search" (Manovich 1999, 81). In his text, the concepts of collection and database are synonymous and interchangeable. He gives the following examples of databases: multimedia encyclopedia, virtual museums, CD-ROMs, considering (from a structural point of view) different modifications of what in this work is understood as a digital archive.

[84] A similar idea, although perhaps in more accurate terms, is suggested by Christian Vandendorpe as he compares the structure of the printed and the digital text. To this end, he introduces the opposition of linearity and tabularity, where "linearity designates a series of elements that follow each other in an inviolable or preestablished order... This concept contrasts with that of tabularity, in which readers can visually access data in the order they choose, identifying sections of interest beforehand, in much the same way as when looking at a painting the eye may contemplate any part" (Vandendorpe 2009, 22).

[85] Manovich's overtly one-sided definition of the narrative as linear and sequential has repeatedly drawn criticism. Jerome McGann, for instance, noted that "'modern age' – presumably, here, the modernist twentieth century – is famous for the inventive ways it fractured and overthrew narrative, especially 'privileged narrative'. But Manovich needs an easy binary to install the progressivist story..." (McGann 2007, 1589).

[86] Manovich views both forms exclusively in terms of structure, and does not consider, for example, the social aspect emphasized by Bruno Strasser: "Databases reflect not only the coming age of modern computer technology to deal with data deluge but, more important, the creation of a new social and moral order with distinct communities that collectively contribute to the production of knowledge" (Strasser 2019, 6)

Drawing heavily on Manovich's analysis, Ed Folsom, in turn, celebrates the emancipative potential and heuristic possibilities of the database. Using *The Walt Whitman Archive*[87] as an example, Folsom argues that the database exempts knowledge (in this case Whitman's poetry and biography) from rigid categories, genres, and hierarchies of narrative, offering instead an experience of the information flow. First, the database makes it possible to discover lost connections. Its fractal and rhizomatic structure provides means for a much more natural and fruitful reflection and exploration of the diverse and non-linear relationships in both Whitman's poetry and biography, than was possible in a narrative. Second, the database allows for accumulating and bringing together a wide variety of materials, facilitating "access, immediacy and the ability to juxtapose items". Third, the database offers an opportunity to transcend the selectivity and partiality of the narrative: it is "the only way to represent the universal … a thousand bricks, all the particulars with none left out" (Folsom 2007, 1575).

Defining the digital collection through the opposition of the narrative, Manovich and Folsom radicalize the distinction between digital and print cultures. In their interpretation, the database stands as a fundamentally democratic form that has the power to discover and produce heterogeneous and manifold relationships between entities. This argument has become a commonplace in the theory of databases, hypertext and digital storytelling[88]. Yet this opposition remains more at the level of theory than of actual practice. For example, Meredith McGill (2007), in her response to Folsom's article, notes that the Whitman archive does not actually get far from print culture, and in fact even adheres to it by reproducing a few specific volumes[89] of and about Whitman[90].

Jerome McGann (2007) offers an even more subtle argument regarding the democratic nature of the database. He observes that the database actually requires a far more well-defined categorical grid than the tangible card system of a paper archive. The paper-based system of the traditional archive is much more flexible simply because of its historicity: it reflects the traces of numerous edits, changes, additions, along with the comments and observations left by researchers in the course of their investigations. Unlike the paper archive, the database requires preset categorizations, and therefore can hardly liberate knowledge from the oppression of categories and genres, as promised by Folsom.

---

87 https://whitmanarchive.org

88 See for example Page, Thomas 2011.

89 Digital edition of printed volumes of personal papers is a fairly common practice: see e.g., *The Collected papers of Albert Einstein* project [https://einsteinpapers.press.princeton.edu]

90 As McGill puts it, Whitman's archive only "gestures toward the world outside Whitman's writing but zigs and zags mostly within itself" (McGill 2007, 1594).

2 Communicative argument

Whereas Manovich and Folsom formulate structural differences between the archive and the narrative, Franco Moretti draws attention to the pragmatic or communicative aspect. As he succinctly puts it in one of his interviews, "a text always speaks to us; an archive doesn't" (Hackler, Kirsten 2016, 11)[91]. Drawing on Moretti's point, one can say that the archive not only does not address the reader, but on its own (in the absence of a research question) it communicates nothing. If the coherence of a text implies a message supported by both the intention of the authorial instance and the intention of the reader, the archive does not form one. Not only the message is missing in the archive, but also the very situation of the utterance: formally, no one speaks and no one listens. Therefore, all those markers that are traditionally used in the narrative to situate and frame the position of the author and the reader[92] are absent in the archive. The user's path through the digital archive is not set by the laws of reading and exegesis, but rather by the laws of navigation or surfing, described by Christian Vandendorpe as "moving in an uncharted environment with no stable landmarks, no precisely plotted routes" (2009, 116). Perhaps for the first time, when encountering the past, we have no guide (a narrator) or stable context (a sequence of movie frames, museum halls, or archive boxes, where each previous instance contextualizes the following one). Instead, we get in touch with the past through bizarre ornaments of images, texts, sounds, a jumble of fragments that may have their place in different stories, but in the absence of a single frame of reference are nothing more than noise.

According to Moretti and the "communicative paradigm," the production of meaning in the digital archive can only be accomplished by the reader having a "good question to ask these archives" (Hackler, Kirsten 2016, 11): only then, through the prism of that question, can the archive have a "message", being a response to that question. Such a perspective makes it clear that the logic of traditional hermeneutics and narrative theory does not work well in relation to the archive, exposing all the aspects that the archive *does not* have. The communicative paradigm treats the archive as if, in confronting the narrative, it becomes degree

---

[91] Arlette Farge makes a similar point with regard to the traditional, pre-digital archive: "il [le texte] est charge d'intention; la plus simple et la plus évidente étant d'être lue par les autres. Rien à voir avec l'archive; trace brute de vies qui ne demandaient aucunement à se raconter ainsi…" (Farge 1997 [1989], 12).

[92] Roman Jakobson, after Otto Jespersen, called those markers, referring to the event of the statement itself, shifters (e.g. *me*, *you*, *here*, *tomorrow*, etc.).

zero of representation. It seems that it overlooks the politics and mediality proper to the archive, namely certain order and regime of encountering the past that it imposes.

3 Media-technological argument

Apart from the structural and pragmatic justifications, there is a third, technological, argument advocated by Wolfgang Ernst in his program of media archaeology. In his definition of digital and multimedia collections, Ernst appeals to a profound technological layer: in his view, the core of a digital archive is constituted by mathematical operations – as he himself formulates it, "counting, rather than recounting" (Ernst 2013, 71). The technological *arché* of the digital archive is indifferent to the content, semantics, interfaces, its very essence being calculations, algorithms, and transmission protocols. For Ernst, narrativity is in line with traditional archives, which are deeply rooted in the age of print and inseparable from reading practices. By contrast, the digital archive, grounded in operative mathematics, liberates history from the oppression of the scriptural regime, rhetorical figures, narrativity, semantics[93] and the primacy of hermeneutic understanding. The *dispositif* of the digital archive as a medium, in other words, implies that the archive is not read and understood, like a text in the hermeneutic tradition, but rather is uncovered through data processing.

Although Ernst concurs with Manovich on the main point – the very opposition of the digital archive to the narrative, as well as the need to develop a new approach to describe it – their arguments and conclusions are very far apart. For Manovich, the database is still a genre, a form, whose syntactics and semantics can be explored by developing a grammar of the 'language of [this] new media'. For Ernst, the digital archive ceases to be a semantic form and turns into a technological *dispositif* in which there is no place for any ('human-oriented') meanings or values, but only for the message of the machines, which we can try to listen to. As he puts it, "Media archeology concentrates on the non-discursive elements in dealing with the past: not on speakers but rather on the agency of the machine" (Ernst 2013, 45).

Withdrawing the entire semantic and discursive dimension, Ernst proposes to deal not with the carefully stored heritage, content or message of the archive, but with what he calls archival noise: "a *mémoire involontaire* of past acoustic, not intended for tradition: a noisy memory, inaccessible to the alphabetic or other symbolic recording, added by the channel of

---

93 In this conceptualization of the digital as anti-semantic, Ernst echoes the ideas of Bruno Bachimont, who defines digitization as: a "paradoxical operation of stripping the entity of its semantics, of abstracting it ... from its semiotic environment to turn it into a manipulable entity" (Bachimont 2020, 56; my translation from French).

transmission – the proverbial "medium" in Claude Shannon's *Theory of Communication*" (Ernst 2013, 175).

In traditional mathematical theory of communication, noise is understood in opposition to information: noise is both what disturbs, interferes with the signal, distorts information and what is not encoded as information[94]. To put it differently, noise is both interference and *non-information*. In his short essay "Message et Bruit", to which Ernst draws attention, Michel Foucault shows that the message itself is only possible against a background of noise (Foucault 1994 [1966]). Only in the presence of noise and certain conventions for deciphering it can a message be transmitted. Furthermore, Foucault suggests that as conventions or systems of knowledge change, the noise could be turned/reinterpreted into a message (and vice versa). To give Foucault's own example: Freud changed the conventions of medicine by proposing to interpret (and thereby transforming into a message) the verbal manifestations of illness, which had previously been regarded as noise.

Based on this peculiar dialectic of noise and message, and in accordance with his definition of digital archives, Ernst designs his methodological program of media archaeology. In it, he proposes to shift attention from archival semantics to the materiality of media itself: literally, to listen to the noise of old machines (e.g., the scratch of the wax cylinder on old phonograph recordings) and to reinterpret this noise as a message to be studied by the digital archive researcher (media archaeologist). According to Ernst, shifting the focus to noise draws attention to the technology itself, its materiality, tactility, and mediality and thereby it provides a way to discover "not only the memory of cultural semantics but past technical knowledge as well, a kind of frozen media knowledge embodied in engineering and waiting to be revealed by media-archeological consciousness" (Ernst 2013, 182).

Ernst's media archaeology program is a very influential way of thinking about digital archives. However, despite perhaps a fair appeal to the "technological" *arché* of the digital archive, it (dramatically) disregards and neglects the historical experience of the user. The assumption that we can explicate a-semantic "technical knowledge" does not at all contradict the fact that the archive also preserves near-surface "semantic knowledge" which offers the user a certain experience of the past.

The three arguments presented above offer three very different ways of addressing the issue of semanticity of the digital archive. The versions range from the idea that the digital archive brings new opportunities for the production of meaning (Manovich and especially

[94] On the conceptualization of noise, see also Johnston 2010.

Folsom) to the complete renunciation of its semantics in favor of non-symbolic and non-semantic noise (Ernst); from the radical individualization of the reader's experience and path (communicative paradigm: meaning depends on the reader's request) to negation of the reader's optics in favor of the optics of the machine (Ernst).

All three of the approaches described, however, elaborate the theoretical (or ideal-typical) models of the digital archive and do not engage with actual collections. Therefore, I find it reasonable to verify and refine the theories by looking at how existing archives (of science) are organized, how they function and give meaning to the artifacts exhibited. This further discussion is intended as both an illustration and clarification of some of the theses discussed above.

2.1.2. Situating object within the digital collection

My starting point will be one particular archive, the "Discovery and Early Development of Insulin" collection at the University of Toronto[95]. Launched in 2003 and updated in 2017, the collection draws together materials across a number of holdings and repositories of the University of Toronto and Balmer Neilly Library. The collection focuses on the early history of the discovery and use of insulin from 1920 to 1925. As such, it represents a case of an uncommonly event-driven archive, a collection meant *to tell the story* of one important discovery. In its task, it hence comes as close as possible to the narrative, which is why I chose it as the object of analysis.

The episode "narrated" by the archive is set by its title itself. The discovery of insulin is a conventional "great scientific event", requiring no special interpretation, easily and naturally embedded in the history of scientific progress and the accumulation of innovation. Yet one fails to grasp and capture the very event of the discovery within the digital collection: it is scattered between protocols of experiments on dogs, newspaper clippings, notebooks, correspondence with patients, notes on insulin injections, and tables with blood sugar readings. A plethora of archival artifacts prevents the event from being fixed and localized in time and space. Indeed, what form of documentation would allow a discovery to be considered accomplished: the first public announcement of the discovery of insulin[96], a measurement of blood sugar drop in the first patient[97], early experimental data from laboratory notebooks

95 https://insulin.library.utoronto.ca
96 https://insulin.library.utoronto.ca/islandora/object/insulin%3AT10010
97 https://insulin.library.utoronto.ca/islandora/object/insulin%3AM10015

showing promising results in dog 35[98], or maybe the US patent for the extraction method[99]? An event that can be recounted in a single phrase (say, the Nobel Prize Committee's prize-giving formula "for the discovery of insulin") is here dispersed into 9,000 separate fragments – pieces of material evidence that maintain, reinforce and document that single fact.

Narrating the story of insulin discovery inevitably involves a procedure of selection: only a tiny fraction of this massive body of evidence can be used. Recounting the story as "a tale of monstrous egos, toxic career rivalries and injustices"[100], "The Conversation" magazine, for example, selects a total of 11 objects from the archive to document and support it with archival evidence (there would probably be more references to documents in an academic historical study, but hardly 9000!).

Within the digital collection, the main way to navigate and orient oneself in these thousands of artifacts and fragments is through subject categories that aggregate individual (thematic) pieces of the story. The "Nobel Prize" category, for instance, picks out and draws together all the items that are relevant to that individual episode within a "big event" of insulin discovery*:* numerous newspaper clippings, remnants of countless communications, seating plans for the banquet, letters announcing banquet times and those thanking relatives for the invitations, congratulations full of courtesan pleasantries, convocations, and announcements (Fig. 2.1.1, on the left).

The resulting panorama draws together artifacts across collections and across narratives. Moreover, upon closer look, it incorporates and exposes some *portions/ fragments* of the artifacts specifically related to the Nobel Prize, and places them on the same level and in equal position with the "stand-alone" objects. So, for example, it includes a newspaper clipping about the Nobel Prize presentation (Fig. 2.1.1, on the right). This clipping originally forms part of a scrapbook compiled by Frederick Banting, one of the principal insulin discoverers. In the digital collection, however, the clipping appears as a stand-alone separate exhibit that can be juxtaposed with other artifacts (or portions of artifacts). Within the Nobel

98 https://insulin.library.utoronto.ca/islandora/object/insulin%3AN10013

99 https://insulin.library.utoronto.ca/islandora/object/insulin%3AQ10017

100 https://theconversation.com/the-discovery-of-insulin-a-story-of-monstrous-egos-and-toxic-rivalries-172820
The outline of the story is as follows: the Nobel prize was divided between Frederick Banting, the main author of the idea, and professor John MacLeod, his supervisor, who provided him with a laboratory, but was not involved in the experimental work. Banting was "furious" that Charles Best, his assistant, working alongside him on the experiments, was not awarded, while his supervisor was. Banting subsequently shared his part of the award with Best, and MacLeod, in return, divided his part with another member of the team, James Collip. The entire controversy is discussed in detail in Michael Bliss's monograph (1982), which is considered the most authoritative academic version of the story.

Prize category, it acquires equal status with, for example, the gold Nobel Medal awarded to Banting and engraved with his name – perhaps the most precious and tangible of the traces, which proves and certifies the very fact of the award[101].

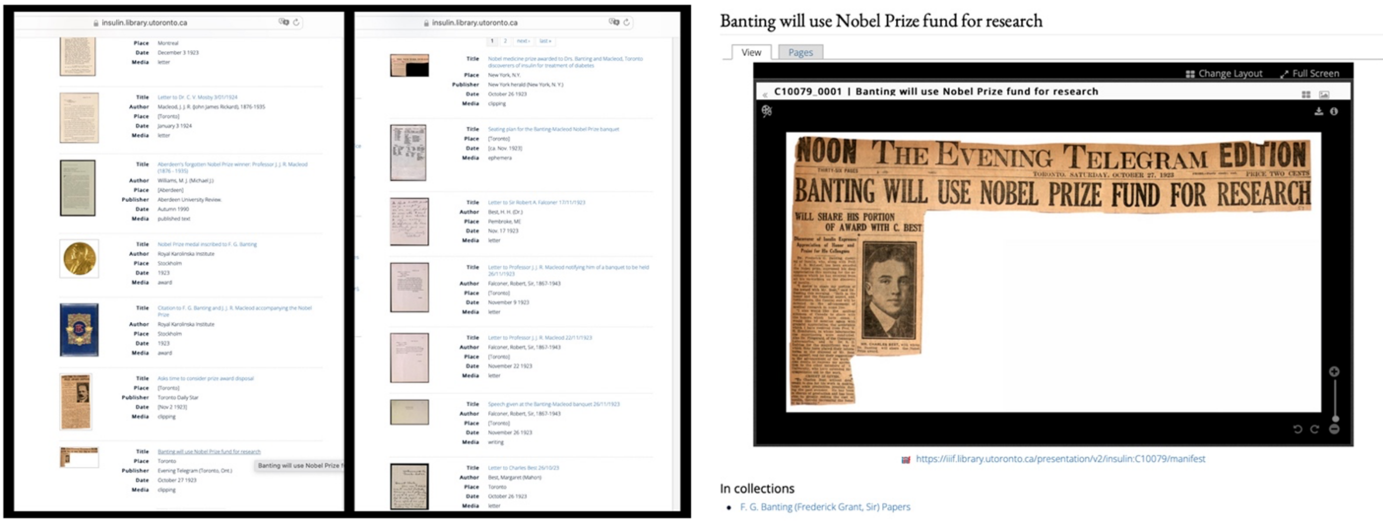


Figure 2.1.1. On the left: objects within the "Nobel prizes" category. On the right: A newspaper clipping from Banting's scrapbook represented as a separate item. "Discovery and Early Development of Insulin" collection, University of Toronto, [https://insulin.library.utoronto.ca/islandora/object/insulin%3AC10079].

In the traditional archive, this newspaper clipping does not stand on its own, but only within the context of Banting's collections of newspaper clippings, letters, photographs and other memorabilia (Fig. 2.1.2).

Each scrapbook constitutes a personal narrative, a special form of memory, in which every clipping and every piece of paper serves as a part of Banting's scientific and personal biography, as he himself had designed and imagined it. In the digital collection, this entire memory collage has been broken down into separate fragments used to exemplify archival subject categories. All of these fragments and cuttings were assigned their own titles, identification numbers and descriptions, making each of them appear understandable in itself.

[101] https://insulin.library.utoronto.ca/islandora/object/insulin%3AA10012

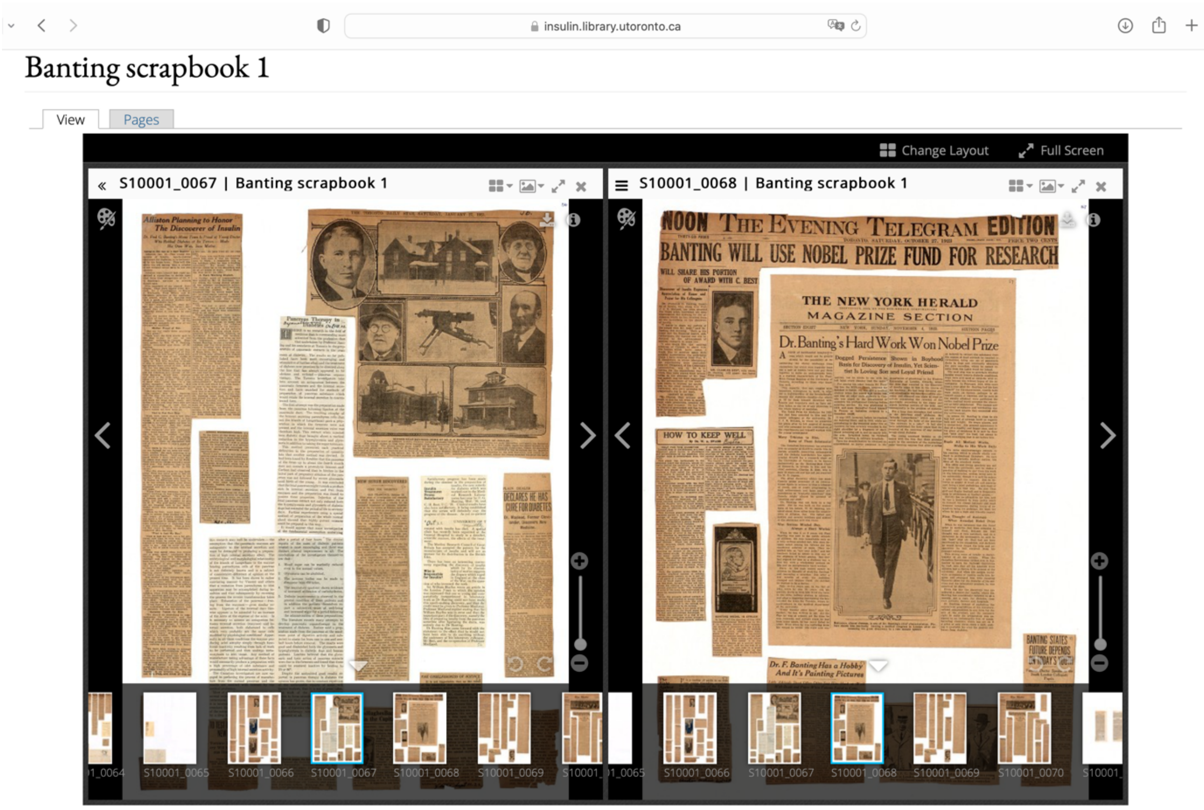


Figure 2.1.2. The newspaper clipping within Banting's scrapbook [https://insulin.library.utoronto.ca/islandora/object/insulin%3AS10001_0068]

In the traditional archive, it would be inconceivable to extract an item from Banting's handmade scrapbook and insert it into some other one. As discussed in the previous part, the traditional archive preserves the integrity of a thing, the order and conditions of its production and storage, which are prioritized above the semantics of objects. In the traditional archive, there is a long way to go in finding this clipping: one needs to follow the path from the *Banting, F. G. Papers* collection to the *Biographical materials* series, on to the *Scrapbooks, 1890s-1941* subseries, and further to a certain folder of a certain box. The search takes the form of an immersion into the archive, where each successive iteration further contextualises, narrows and specifies the search. The traditional archive, in other words, is to be structured topographically rather than semantically: it reconstructs not the values but the conditions of its production (media, technology, bureaucratic practices). In Sven Spieker's observation, its structure is similar to an archaeological site where the place of discovery is the main defining feature of an item[102]. This desire to put things in their place is reflected in the very multi-level archival structure divided by series and sub-series, fonds and sub-fonds.

102 Spieker also draws an interesting analogy between the "topography-centered" traditional archive and 19th-century science: he cites the cell theory of Rudolf Virchow, one of the pioneers of anatomical pathology, who "treated pathological tissue in exactly the way that an archaeologist treats a fragment he finds in the ground or the way a nineteenth- century philologist treated words: as discrete, isolated

Jacques Derrida treats this attachment to territory as a precondition for the archive's power: "Even in their guardianship or their hermeneutic tradition, the archives could do neither without substrate nor without residence. It is thus, in this *domiciliation*, in this house arrest, that archives take place" (Derrida 1995, 2). The digitization of the archive entails its homelessness, the absence of residence: any artifact, or even a fragment of an artifact, can be juxtaposed with another one and accessed outside the logic of the original order. Whereas the traditional archive preserves the integrity of a thing, the digital archive fragments that integrity into many smaller units, which can then be combined and recombined in any possible order. As we have seen, Manovich and Folsom recognize the emancipatory potential of this new order that is to bring about new meanings and new relations.

But how is the artifact situated in the digital collections if its residence is no longer meaningful? And how it is to be read out of the "topological" context maintained by the traditional archive? Let us return briefly to the example of the newspaper clipping. Within the digital collection the user is instructed to read it as evidence of the Nobel Prize presentation. In addition to semantic / thematic categories, the user is also guided by the archival description / metadata of a given newspaper clipping (Fig. 2.1.3).

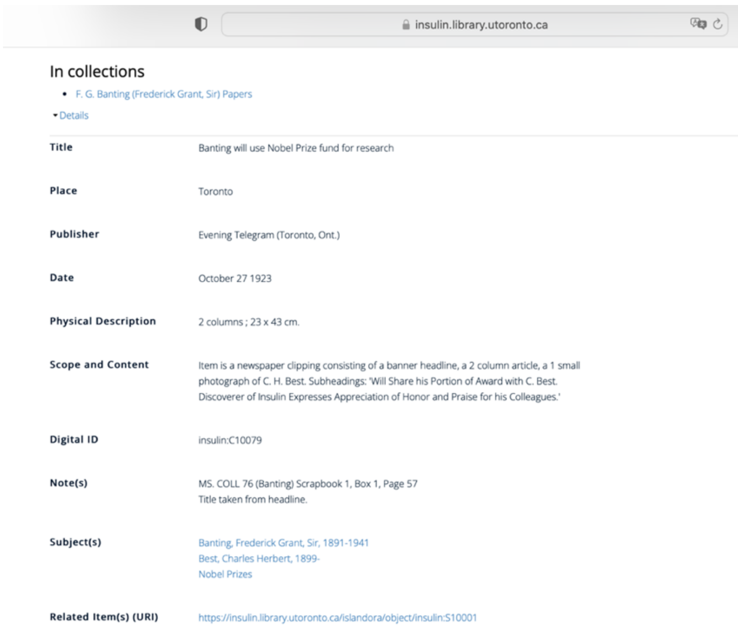


Figure 2.1.3. The item-level description of the clipping, [https://insulin.library.utoronto.ca/islandora/object/insulin%3AC10079]

pieces of evidence that can be understood only in the context of the place (and the time) where they were detected"(Spieker, 2017, 19).

This description operates with numbers and matters of fact. It captures the source, place and date of the publication, provides the physical description of the clipping, and a specification of its "content": the fact that it consists of "a banner headline, a 2 column article, a 1 small photograph of C. H. Best". The descriptive metadata performs two interrelated tasks: first, it stresses the authenticity, credibility, facticity of the clipping (describing its source and material characteristics), and second, it locates and identifies it in the (physical) archive. Taken together, the semantic categories and item-level description, therefore, signal to the user that the newspaper clipping constitutes an authentic fragment of the past, which bears witness to the awarding of the Nobel Prize.

Now, let me recall that this newspaper clipping in fact forms part of a coherent narrative, created, fabricated, designed by Banting. It is the result of selection and craft, cutting, gluing; a complex object in which, for a historian, the clipping lines may be more instructive than the text of the article. In other words, this clipping acquires meaning precisely as part of the whole (narrative). In the digital collection, by contrast, it is represented as an isolated fragment of the past, as a simple, accidentally surviving and preserved testimony (about the Nobel Prize, rather than Banting's mnemonic strategies!). As such, it is revealing that the title assigned to this clipping as a separate and self-valuable artifact simply repeats the title of the represented article.

The clipping thus is converted from a complex sign into an emanation of the past. The collection describes items as an index or trace of reality (and it makes no difference of what order of reality). The archive frames it as a matter of fact, a 'neutral', objective fragment of reality, to use Roland Barthes' formulation, as "the pure and simple 'representation' of the 'real,' the naked relation of 'what is' (or has been)", which requires no interpretation and "appears as a resistance to meaning" (Barthes 1989 [1968], 146). In this instance, we are thus dealing with the *naturalization* of an object within the digital archive as opposed to its *semantization*. Taken out of its context of production (as, in fact, any other context), the artifact is characterized by means of an "objective" archival record that offers no new context, but attests to its authenticity, facticity and materiality[103].

The newspaper clipping interpreted outside the context of a scrapbook is a rather extreme example. However, this very form of describing an artifact and giving meaning to it constitutes not an exceptional case, but rather a common archival practice. In order to observe

[103] As discussed in the previous part (Chapter 1.4), hereinafter I do not ascribe certain motives or decisions to individual actors or systems. Rather, I outline certain patterns of archival representations and their *effects* (stipulating that the underlying causes of these effects may differ from case to case).

overall tendencies, I traced how item descriptions are organized in the 118 collections from the corpus. For each of the collections, I indicated whether the following forms of description were employed:

- *formal descriptions*, such as descriptions of the size of the artifact, the material it is made of, its genre;
- *dates*;
- *content descriptions*, such as listing the names of people depicted in a photograph or outlining the content of a piece of text;
- *historical background* – by whom, when and under what circumstances the object was created or used, the object's provenance or a biographical note;
- *context* – the broader historical, cultural or scientific context of an artifact, including the context of other related artifacts;
- *references*, such as bibliography or references to some additional materials.

The first three points represent factual information, that which allows to certify the authenticity of the object, its material characteristics. The last three points rather put the object into context, or open it for (further) interpretation. The heat map (Fig.2.1.4) shows the overall pattern of how the item-level descriptions in the collections are organized[104].

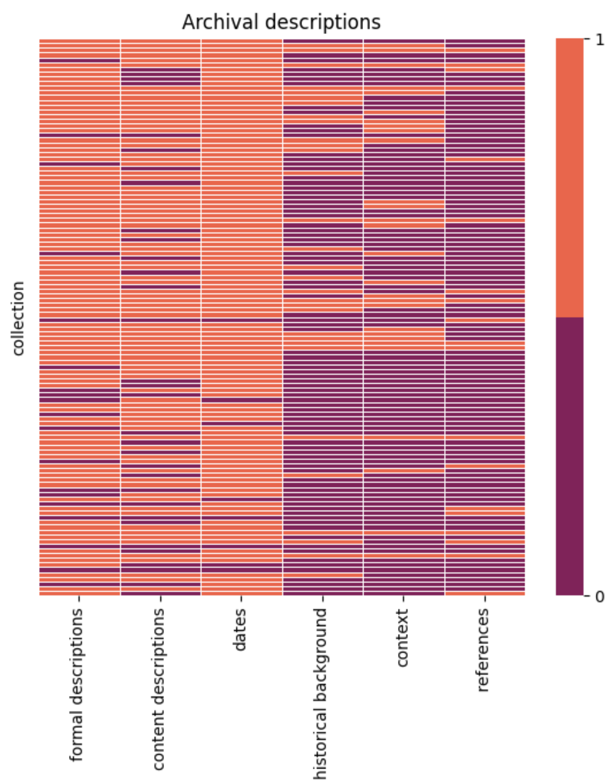


Figure 2.1.4. The heat map represents the way descriptions are organized in the collections. Each line represents a collection. Orange color indicates the presence of some form of description, violet color means its absence.

104 For the data underlying this visualization see Table 2. Item-level descriptions. [https://zenodo.org/records/10548106]

As seen with the naked eye, factual descriptions strongly dominate over "semantizing" and contextualizing ones. In other words, the common tendency is as follows: the digital archives of science are more likely to certify the factuality, authenticity of an object, to frame it as a matter of fact, rather than to provide some context of the object[105]. We thus observe quite a paradoxical structural effect: while the digital archive, as a medium, is devoid of attachment to place and to material objects, yet the adopted practices of archival inventory tend to position artifacts in terms of their materiality and authenticity.

Commenting on the techniques and modes of object description, literary critic Susan Stewart once remarked that they reflect the meaning-making practices adopted within a culture:

> What does it mean to describe something? Descriptions must rely upon an economy of significance which is present in all of culture's representational forms, an economy which is shaped by generic conventions and not by aspects of material world itself (Stewart 1993, 26).

The "generic conventions" we observe in the "digital" descriptions of scientific objects can be traced back to the notions and practices of traditional archives that took shape in the 19th century. These are the practices aimed at ensuring that the object in question is true and authentic.

Going back to the three arguments on the digital archive, there are some adjustments to be made. In setting their models of the digital archive in opposition to both narrative and physical collections, they outline the radically new configuration put forward by the digital archive. In practice, however, what one finds is that (at least some) traditional archival discourses, practices, and standards are transposed and embedded into the digital archive, no matter how new this media is.

[105] These quantitative findings are largely consistent with Jeffrey Schnapp's observations from his analysis of museum online databases (Schnapp, 2018).

2.2. Visibility in/through digital collections

> Works of art will acquire a kind of ubiquity… They will not merely exist in themselves but will exist wherever someone with a certain apparatus happens to be…Just as water, gas, electricity are brought into our houses from far off to satisfy our needs in response to a minimal effort, so we shall be supplied with visual and auditory images, which will appear and disappear at the simple movement of the hand, hardly more than a sign… I don't know if a philosopher has ever dreamed of a company engaged in the home delivery of Sensory Reality.
>
> Paul Valéry

Taking Valéry's thought further: has a historian ever dreamed of having fragments of the past themselves coming to their doorstep? If so, the Science Museum in London recently brought that dream to fruition by developing a special browser extension that every day shows the user a random object from the museum's digital collection[106] (Fig. 2.2.1).

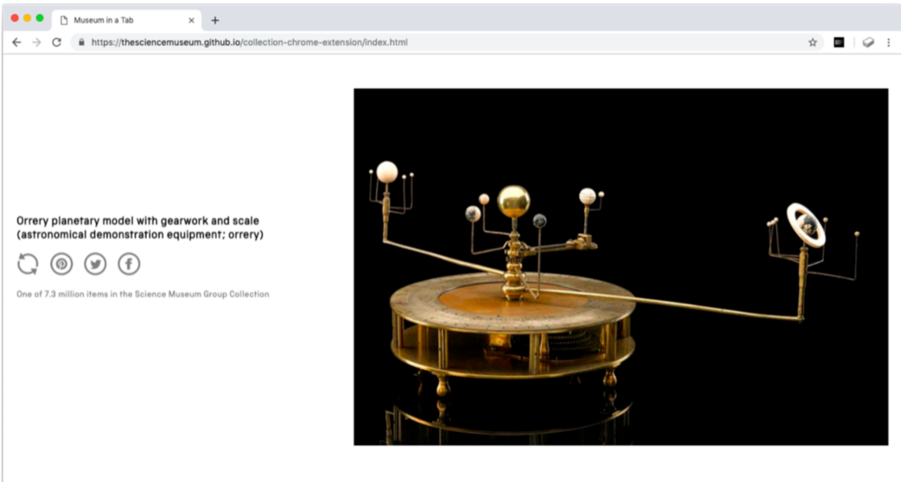


Figure 2.2.1. "Museum in a Tab" project. The Science Museum in London.

Fragments of the past no longer reside in the closed sanctuary of the archive; instead, they are exposed to our gaze. In Walter Benjamin's terms (1936), this means that the cult value of archival artifacts is finally and irrevocably replaced by an exposition value. Archival objects become not just accessible (as is commonly noted in discussions of digital collections), but are made *visible* and *ubiquitous* – to the point where they can be delivered automatically to the computer screen, even without any prior request on the part of the historian.

This chapter explores the visual *dispositif* of the digital archive: that is, how do historical artifacts come to be *visible* in digital collections. How is the graphical interface of digital collections organized and what are its affordances? What interpretations does it make

[106] https://lab.sciencemuseum.org.uk/museum-in-a-tab-chrome-extension-4251234a337e

possible (Drucker 2014; Whitelaw 2015)? How does the visual experience of coming into contact with the (digital) residues of science work? In answering these questions, I will successively examine two viewing modes through which the user encounters the exhibited objects: the panoramic view of items (*digital ornamentality*) and the one-on-one encounter with the object (*zooming in*).

2.2.1. Digital ornamentality and the economy of vision

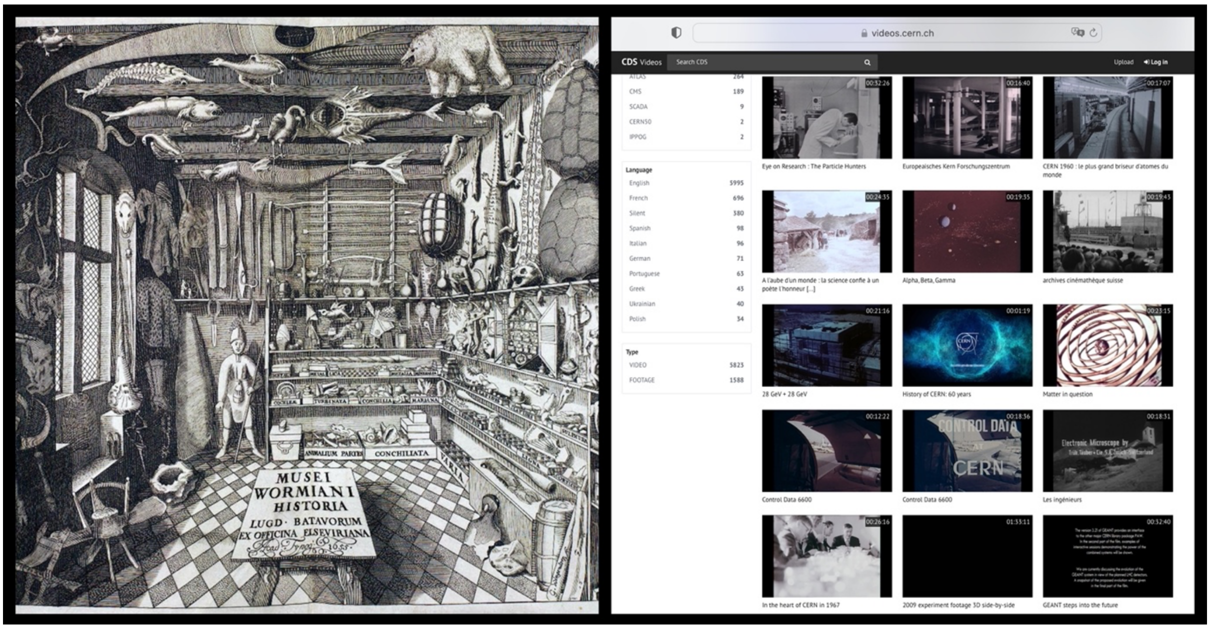


Figure 2.2.2. Ole Worm's Wunderkammer and the interface of the CERN video collection

Consider these two figures (Fig. 2.2.2) representing two knowledge landscapes, separated by almost five centuries. The first is *Wunderkammer*, a cabinet of wonders – a theatre and laboratory of the mid-17th century, established by Ole Worm in Copenhagen. The second is the video collection of CERN, the world's largest nuclear physics laboratory. The first one was set up towards the end of the Scientific Revolution, the second one stands in the midst of the era of Big Data, when science is being tirelessly memorialised and becomes interested in looking at itself.

Both constitute *theatrum scientiarum*, exhibiting carefully selected "scientific objects" – and yet objects of wonder and desire – for the general public. The very idea of scientificity and the public's interest have drastically altered over the last five centuries: from "crumbling shells, clumps of madrepores, coral branches, miniature busts, Chinese porcelain teapots, small medals, intaglio gems, pottery shards" (Stafford 1994: 238) to the animation of the four-top-quark event, presentations at the First World Wide Web Conference, installation of the last piece of the LHC beampipe at ATLAS, and footages of the MoEDAL experiment.

The ways in which these odd objects are arranged and displayed have also changed quite dramatically. The cabinet of curiosities arranges objects in such a way as to together they reflect a unified world order, where all the things resonate and interact with each other. CERN unfolds the overview of objects arranged in an ideal geometric grid, along with the categories to sort and re-organize them. By placing the two interfaces next to each other, I seek to defamiliarize the visual *dispositif* of digital collection we are used to. How does the digital archive locate and *negate* scientific artifacts? What modes of gazing and viewing does it assume?

In most collections from the corpus, the entry point is precisely the overview of objects[107] that we observed with the CERN video collection. This presentation mode offers a panorama of the available objects, from which users can pick the objects they are interested in and click on them to take a closer look (Fig. 2.2.3).

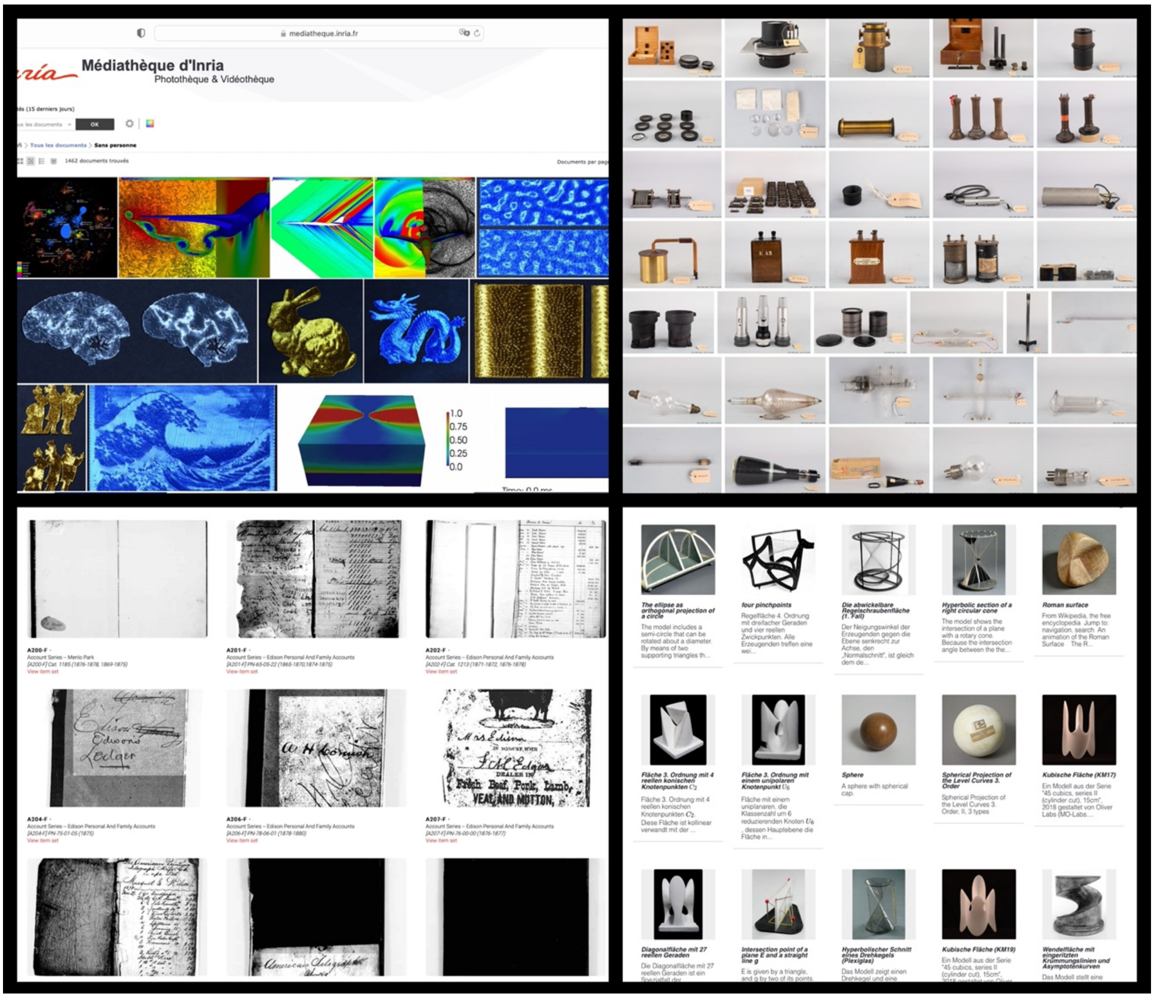


Figure 2.2.3. The ornaments of the INRIA collection (top left), ETH Zurich collection of scientific instruments (top right), collection of Thomas Edison Papers (bottom left) and Digitales Archiv Mathematischer Modelle (bottom right).

[107] For some alternative and experimental examples of interfaces and visualizations of collections, see Whitelaw 2015; Kräutli 2016 ; Bludau et al. 2021.

As opposed to the keyword search, this display mode invites visual exploration, browsing as opposed to querying, and therefore represents an example of what Mitchell Whitelaw called a "generous interface", showcasing "scale and richness" of the collections (Whitelaw 2015). No wonder it is often referred to as an "image gallery", transparently alluding to the experiences of art galleries. This analogy, though, seems somewhat misleading to me. In the art gallery the spectator moves (e.g., through rooms) sequentially looking at the paintings, one after another, and under no circumstances can cover all the pieces at once with his gaze. The digital archive, on the contrary, arranges objects in such a way that the user's eye can cover as many of them as possible, and offers not a sequential journey but rather a holistic and synchronous panorama of items.

I will refer to this representation mode as the *digital ornamentality.* "Ornamentality" draws attention to the decorative arrangement, the fact that the items are arranged in regular and standard shapes and positioned with a certain rhythmic pattern. "Digital" emphasizes that the two basic procedures of ornamentation – framing and filling – are automatically performed by means of computer. Modern platforms for creating digital collections (such as Omeka.org) offer ready-made versions of ornaments – patterns of artifact arrangements – that are then filled with specific images as desired by the user. When designing an exhibition in one of such repositories, ornamentality is essentially impossible to avoid; it appears to be the principal mode of organizing images in the digital archive, prescribed and predetermined by the very infrastructures for exhibiting digital collections.

The concept refers not only to a particular way of decoration, but also to the politics of order as articulated in Siegfried Kracauer's "The Mass Ornament" (2005 [1927]). For Kracauer, the ornament comes to be seen as an expression that manifests and carries a certain social and political agenda. Thus, he interprets the rhythmic figures of the "Tiller girls" dancers as a manifestation of a new type of social organization in which community is replaced by masses.

The treatment of the historical artifact in the digital archives fits well with Kracauer's critique. The archival ornament, much like that of the Tiller girls, imposes a certain (decorative) order on the material. In it, the artifacts are subjected to normalization and standardization: they populate identical square shapes arranged in a definite geometric order. By equalizing and aligning objects, the ornament emphasizes the *multitude*, the masses of things. As a distinct and novel object of historical representation, the *multitude* of things challenges, firstly, the uniqueness (or aura) of each individual object, and secondly, the idea of communality or "being-together" of things, its meaningful and concrete relations with other things. This type of display is more akin to online platforms like Amazon than to conventional historical loci such as a museum or an archive (Fig. 2.2.4).

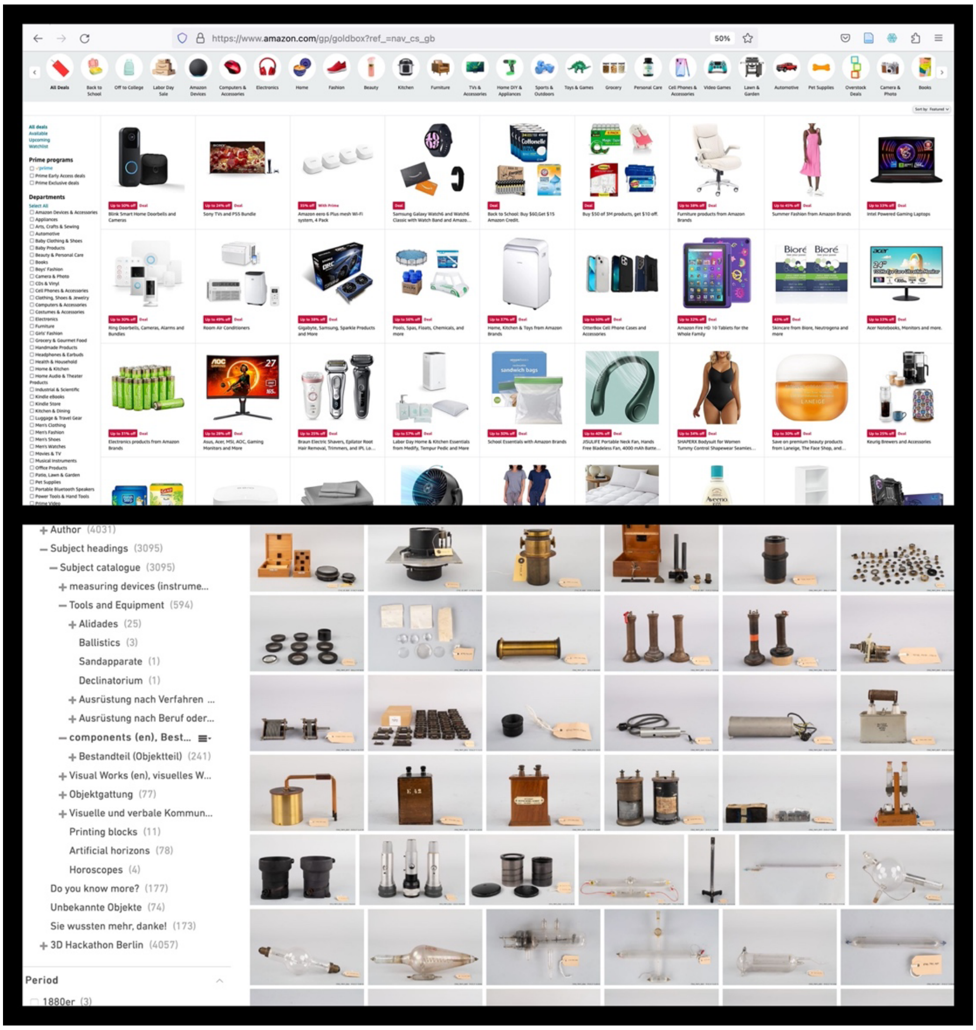


Figure 2.2.4. The Amazon offers page (top) and the ETH Zurich collection (bottom)

Now, what happens when this both decorative and political configuration, the digital ornament, serves as a matrix for the representation of historical artifacts? In terms of historical representation, the most important characteristic of the digital ornament is perhaps the absence of perspective, or in Anne Friedberg's wording, "post-perspectivity". Instead of a singular frame of perspective, the digital ornament implies "the multiplicity of windows within windows, frames within frames, screens within screens" (Friedberg 2009, 1-2).

Stephen Bann pointed out the importance of the concept of perspective for shaping the historical experience. As he noted,

> the conceptualization of the past in terms of perspectivally ordered space is full of consequence for the experience of 'the past'… writing imposes a regime which is comparable to that of the perspectival painting, in that no detail, or object, is accessible in itself, but is simply an element integrated within the simulating space of the perspective (Bann 1990, 116).

In articulating this perspectival view of the past, Bann found it in historical painting, historical narrative, and the history museum, which all framed and participated in a particular experience of coming into contact with the past. This same perspectival approach is evidently shared by the traditional archive, with its metaphor of moving deep into time: the idea that the historian delves into the archive, layer by layer, in order to dig out the truth buried in its depths.

The depth of the physical collection contrasts with the simultaneity of the digital archive. The digital ornament has neither depth nor a single focal point: it brings to the surface and places all available artifacts on the same level (or plane). The ornament displays everything simultaneously (rather than sequentially) substituting the profundity of the traditional archive with flatness, frontality, openness, synchronicity. In this simultaneity, ornament merges, brings together old and new, everyday and outstanding, standard and deviation. Through it, it becomes possible to display all things at once: Babbage analytical engine, vintage cigarette cartons, DNA structure discovered by Crick and Watson, and old Nokia design do not interfere with each other within one ornament of the Science Museum Group (Fig. 2.2.5).

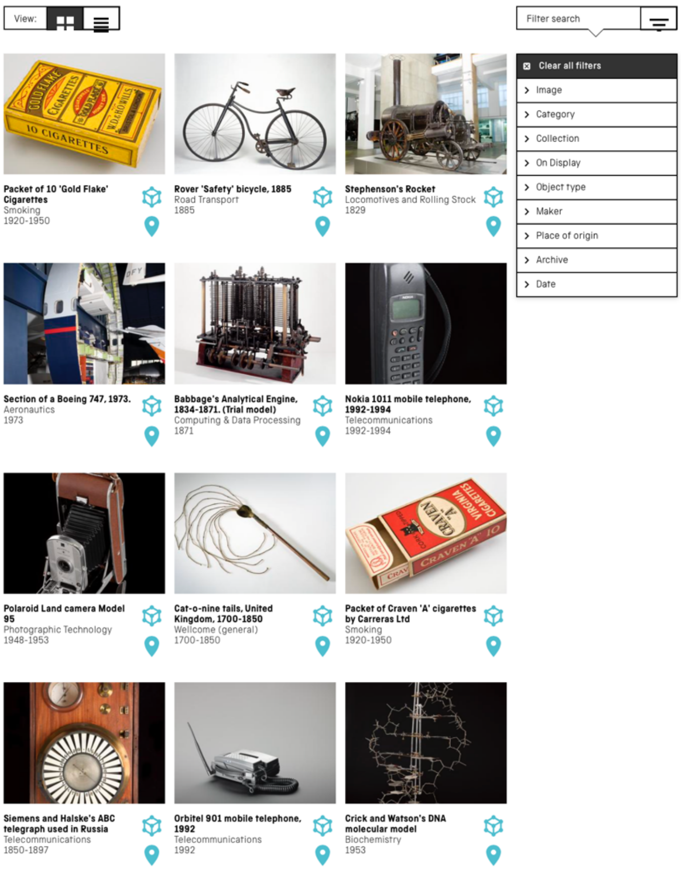


Figure 2.2.5. The ornament of the Science Museum Group collection.

The digital ornament thus puts the spotlight on the multiple rather than the singular and makes it possible to experience everything simultaneously, beyond the flux of history. Thereby it presupposes a particular economy of vision. Art historian Ernst Gombrich terms this effect of the ornament “the etc. principle”:

> Looking at a crowd or a troop of horses we will be less aware of the exact loss of detail because we will tend to expect that the members of this mass will be identical and read them accordingly – finding it difficult, if not impossible, to tell at any point where we see elements and where texture; where we are reading and where we are ‘reading in’ (Gombrich 1994, 99).

Ornament presupposes a gaze that does not examine each thing in isolation but rather observes the multitude, a gaze that does not discern particulars, a gaze in which each individual artifact appears as a part of the overall pattern.

That said, the user does have some control over the digital ornament. Digital archives offer two forms of user interaction with it: sorting and filtering. While users have no control over the pattern of the ornament, they are able regulate its filling: to sort artifacts according to a certain logic or to arrange them in a certain sequence.

To recall, from the standpoint of the structural argument (Manovich/Folsom), this freedom to sort and rearrange artifacts is the main epistemological potential of the digital archive, promising the discovery of unknown and new connections. To put this thesis into practice, I will take a closer look at just one of the possible logics of re-arranging the ornament – namely, the logic of time crucial for the representation of history. To what extent do archives really provide this freedom of arrangement? To what extent can the user take advantage of time markers and categories when interacting with the ornament? At this point, I will draw on the statistics for collections[108] (Fig. 2.2.6).

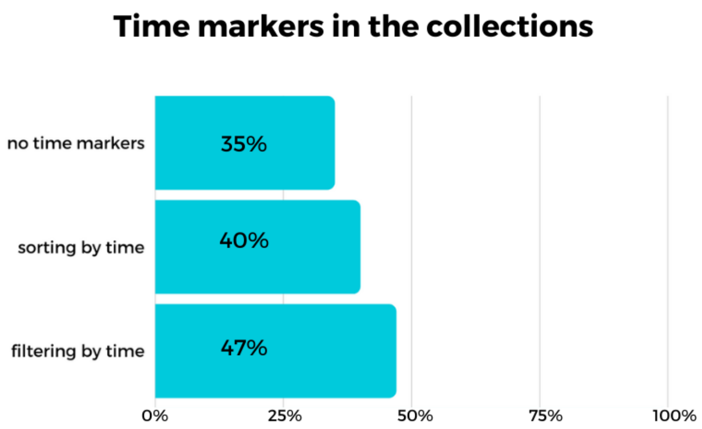


Figure 2.2.6. Time markers in the collections of the corpus.

[108] For the data see Table 3. Time markers.

To begin with a fairly scandalous fact, out of the 118 collections, 42 do not provide any time markers at all. They offer no way to arrange artifacts by time, nor to identify artifacts from a particular time frame. That is, in their structure, about a third of the collections are not subject to historical logic at all.

Further, slightly less than half of the collections (48 out of 118) make it possible to place the artifacts in chronological sequence, from the earliest to the latest (or vice versa). That said, it is only the time of creation/production of an artifact that is taken into account, while all other temporal frames – e.g., the time of its use, acquisition, or archiving – are disregarded. This often results in quite comical effects. For example, in the Sigmund Freud Papers collection (Library of Congress), a chronological sorting reveals the earliest artifact to be a small Greek statue belonging to Freud, dated 6th century BC. Right after it, with a gap of 25 centuries, come Freud's family papers, dated 1851.

What we are dealing with here are fundamentally different temporal orders: the scale of Freud's life and the scale of the life of things that had their own biography before falling into Freud's hands. Within Freud's collection, the thing is not tied to the temporal context of his life, but instead undergoes a (temporal) decontextualization. In the archive these different time orders are superimposed on the same temporal axis, and eventually it appears that Sigmund Freud's personal collection spans from the 6th century B.C. to 1996, when the first edition of Freud's correspondence[109] was published.

Apart from the chronological sorting, in about half of the collections (56 collections out of 118), time markers can be used as filters, that is, one can transform the ornament to only feature artifacts within a certain time period. Different graphical tools are used for this task: timelines, dynamic date range graphs, date sliders, lists, each having its own affordances and organizing user's relationship with time.

I will take a closer look at one, the most frequent[110], of these tools: the list of dates (Fig. 2.2.7). In this case, time markers (years, days, or centuries) are turned into categories arranged according to their frequency in the collection. The scale of time-categories chosen – centuries, decades, years, days, or even the time of day – defines the time frame and temporal context of the collection in question. Following François Hartog's reflections on historical time (2016), we might ask: how does this particular structure of categories organize our experience of time?

---

109 https://www.loc.gov/item/mss3999000441/

110 It is used in 36 collections out of the total of 56 that provide the possibility of filtering by time.

Through such a list, time is rendered discrete, a-chronological, and a-historical; it appears as a set of individual points in no way related to each other. There is no earlier or later, before or after: one cannot trace whether one thing results from another, or precedes it. Moreover, such a list often combines temporal markers at different scales: it may include, for example, both years and days. This temporal orchestration represents neither continuity, nor duration, nor sequence. Thus, these lists of dates do not, in fact, stand as historical markers. Instead, they are non-temporal, a-historical, no longer distinguishing between the past and the present.

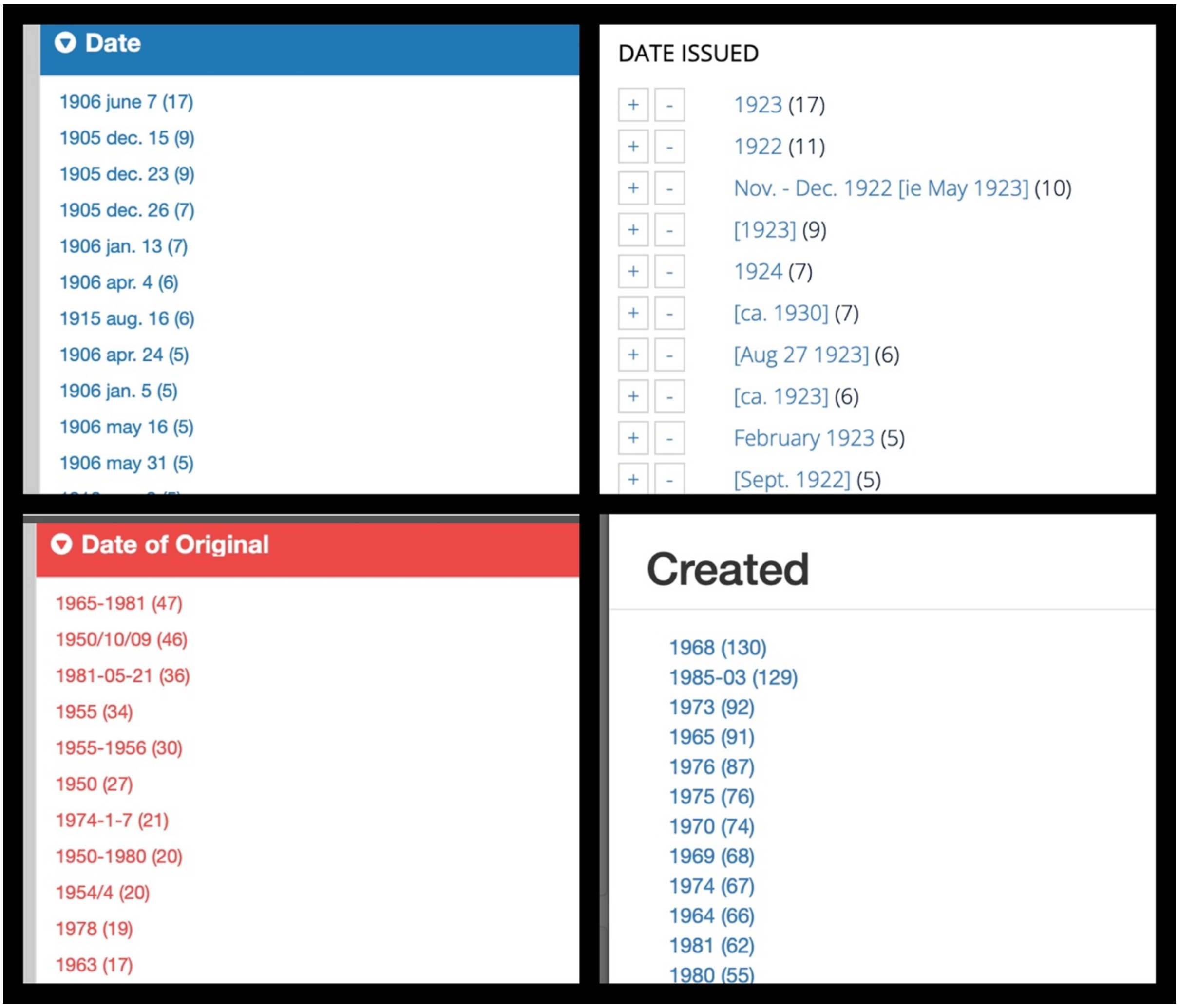


Figure 2.2.7. List of dates in the Frederick Winslow Taylor collection (top left), The Discovery of Insulin Collection (top right), GE Research collection (bottom left) and Charles Babbage Center for the History of Computing (bottom right).

Ornament thus reflects neither continuity nor depth. Instead, it brings forward the omnipresence or ubiquity of digital representations.

In the passage with which this chapter began, Valéry speaks of the ubiquity of images, referring primarily to their accessibility: as he puts it, they will be available to us "at the simple movement of the hand". But, as it turns out, the ubiquity of images also alters our very stance toward the past. The ubiquitous images of (historical) artifacts no longer form the orders of the past we are used to: the older is no longer more distant than the newer. In the digital ornament all is located on one plane, all is equally observable and visible, regardless of the point at which the observer is located.

2.2.2. Zooming in: one-on-one with the thing

In addition to the overarching view of things, the digital archive, of course, involves encountering things one-on-one, putting the focus or *zooming in* on one particular object. This configuration involves a display of one particular artifact, usually occupying the most of the screen, and its metadata (item-level description discussed in the previous chapter).

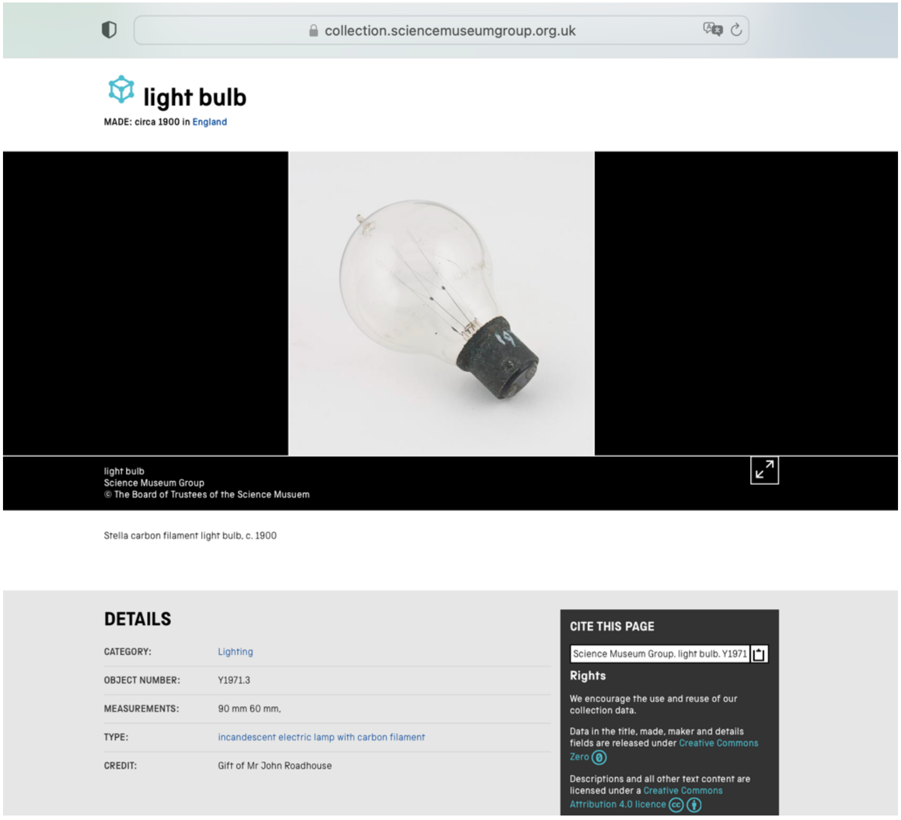


Figure 2.2.8. The example of the "zooming in". The light bulb in the Science Museum Group, [https://collection.sciencemuseumgroup.org.uk/objects/co8404907/light-bulb-incandescent-electric-lamp-with-carbon-filament]

The encounter with an object in the digital archive is typically described as deficient and incomplete compared to the one in the physical archive. In the absence of things, the user deals with their digital representations, which media theorist Vilém Flusser even refers to as the direct opposite of things – *non-things*:

> The electronic pictures on the television screen, the data stored in computers, all the reels of film and microfilm, holograms and programs, are such 'software that any attempt to grasp them

is bound to fail. These non-things are, in the true sense of the expression, 'impossible to get hold of'. They are only open to decoding" (Flusser 1999, 86)

The materiality and tangibility of things from the past has everything to do with the allure of the archives as famously described by Arlette Farge: the "naive but profound feeling of tearing away a veil, of crossing through the opaqueness of knowledge and, as if after a long and uncertain voyage, finally gaining access to the essence of beings and things" (Farge 2013 [1989], 8).

In the digital archive, where the artifact is substituted by its "ubiquitous" representation, the "tactile" and "essential" experience of encountering the thing is respectively superseded by *viewing*[111] experience. Digital humanities scholar Joris van Zundert, for example, describes interacting with medieval manuscripts in some digital environment as "comfortable viewing experience" (van Zundert 2018, 8), "seamless and high-quality viewing experience", noting also that "all that viewing 'power' at one's finger tips is a dream" (ibid., 6). The fingertips as the only point of contact between us and the world of non-things is also noted by Flusser: non-things cannot be possessed, held or grasped by hand.

In the case of scientific objects, which were neither meant to be representations, nor objects of gaze, but means used for science, this material and instrumental dimension is of particular concern. In this regard, the debate on the engagement with the (scientific) artifact refers back to Martin Heidegger's (1962 [1927]) famous distinction between 'presence-at-hand' (*Vorhandenheit*) and 'readiness-to-hand' (*Zuhandenheit*). The first is the relationship, in which the thing appears as an object of contemplation, observation, interpretation, attribution and other forms of detached intellectual study. Heidegger contrasted this mode of encounter with embodied knowledge, 'readiness-to-hand', the knowledge of an instrument in its use. To take his own example, contemplating a hammer and operating it constitute two different encounters with the instrument. What is more, the instrumentality of the hammer only manifests itself through praxis. It can only be grasped and experienced in the very process of *hammering,* which therefore constitutes the principal form of relationship with the thing.

In the digital archive, we are probably farther away than ever from *Zuhandenheit.* Instead, what we experience is *non-presence-at-hand*, an encounter with the representation of a thing in the absence of the thing itself. Inasmuch as Heidegger criticizes the primacy of vision in favor of materiality, the digital collections essentially offer the user only and exclusively

[111] Here and below, I limit myself to the two-dimensional image collections that remain the most common type of digital scientific archives (as exemplified by the corpus). Audio, video archives, as well as 3D digital collections represent more complex cases that are worthy of further exploration.

visual experience. Yet this does not mean that the digital engagement with an object is simply an even more imperfect version of the "presence-at-hand". Rather, as I will argue, it is an independent, new mode of cognition rooted in the nature of the digital.

To that end, I will first discuss the relations between the digital representation of an object and the object itself, or what exactly the user is looking at when zooming in on an object in the digital archive. I will then turn to the forms of engagement with the digital object that archives offer to the user.

Manipulated image

Outlining the nature of photography, Roland Barthes referred to its special "evidential force" (Barthes 2006 [1980], 88-89), its capacity to "authenticate the existence of a certain being" (ibid., 107). By its very essence, photography always points to a referent, a thing, certifies the presence of the depicted object, affirms with certainty that *ça a été* (that-has-been) (ibid., 96). The role of photography as "a certificate of presence" brings about and upholds certain regimes of truth and knowledge that took shape in the 19th century: from evidential practices in court to administrative practices (Tagg 1993).

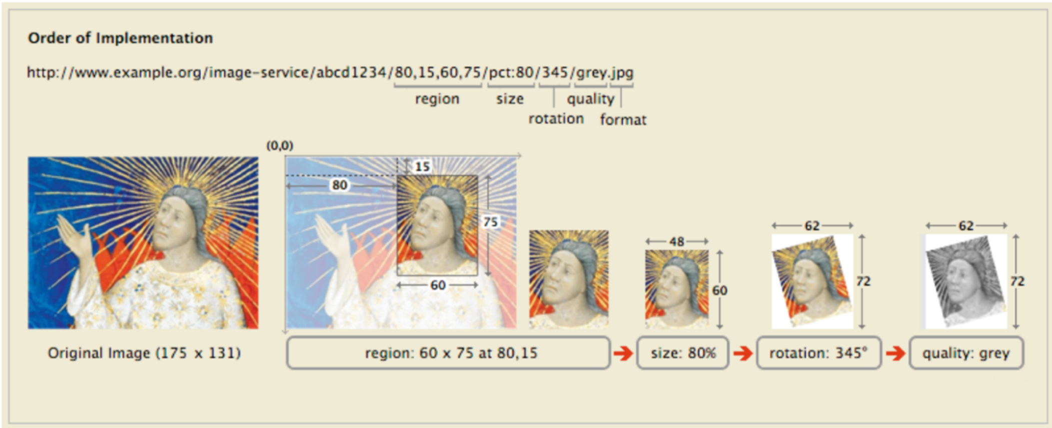
As discussed time and again by the theorists of visual studies, the nature of the digital image marks the end of this mode of representation and regime of truth (Mitchell 1992). Made up of pixels, each of which in isolation depicts and means nothing, the digital image calls into question its very veracity. As Bruno Bachimont remarks, with regard to digital representations, Barthes' old formula should be reformulated as: "ça a été manipulé" (that-has-been-manipulated) (Bachimont 2020, 59). In other words, dealing with a digital image, we are no longer confronted with an emanation of a referent, but rather with a distinct object of a very different nature.

Further, a number of mediating instances intervene to determine how exactly the pixels of the original image are delivered to the user. One such instance is the IIIF[112] standard protocol for image transmission, which is being widely used by archives and collections. The protocol includes the Image and the Presentation APIs, which together define what exactly the user will be seeing through different web-environments [113].

[112] IIIF stands for the International Image Interoperability Framework.

[113] https://iiif.io/get-started/how-iiif-works/. For the complete description of the standard cf. https://iiif.io/api/presentation/2.0/

The Image API defines how image pixels are delivered from the server. It specifies the display parameters of the source image by setting and changing five basic image parameters: region, size, rotation, quality and format. Thus, the image on display in the digital collection may be of a different size or tone than the original image, or it may even be a fragment of some source file inaccessible to the user (Fig. 2.2.9).



The image API controls the form in which an image is delivered to a location on the Web.

Figure 2.2.9. Image transformations within the IIIF Image API. Image source: https://iiif.io/get-started/how-iiif-works/

The image(s) together with the metadata are then displayed in special image viewers compatible with IIIF (some of the examples include *Mirador*, *IIIFViewer*, *OpenSeadragon*). The structure and layout represented through such a viewer is determined by the so-called IIIF manifest (Fig. 2.2.10). As defined in the standard, the manifest is

> the prime unit in IIIF which lists all the information that makes up a IIIF object. It communicates how to display your digital objects, and what information to display about them, including structure, to varying degrees of complexity as determined by the implementer. (For example, if the object is a book of illustrations, where each illustrated page is a canvas, and there is one specific order to the arrangement of those pages)[114].

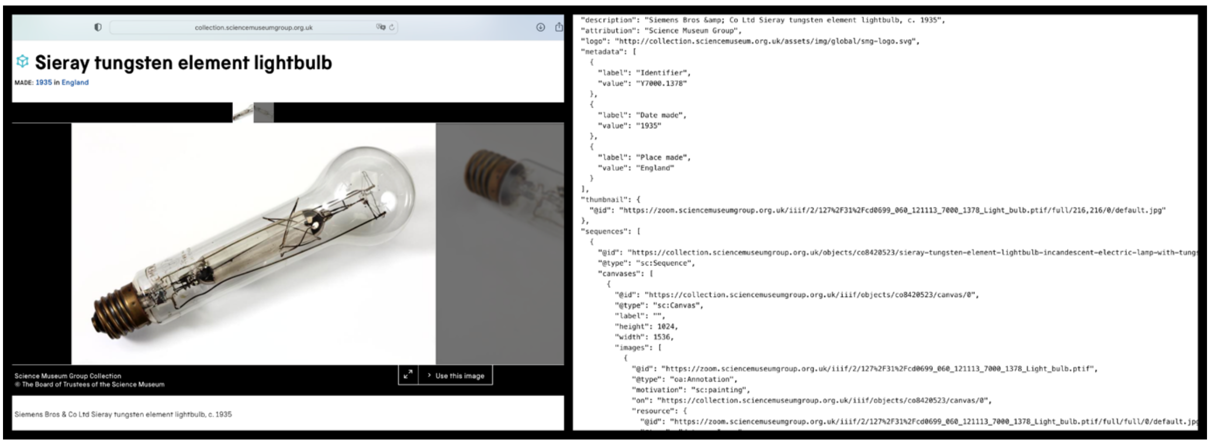


Figure 2.2.10. Another lightbulb from the Science Museum Group collections and its manifest (on the right).
[https://collection.sciencemuseumgroup.org.uk/iiif/objects/co8420523]

[114] https://iiif.io/get-started/how-iiif-works/

The manifest thus contains a reference to the image, as well as its metadata, along with instructions for displaying the image that specify exactly how the object will be rendered: what part of the image we will see, in what sequence, at what resolution. Thus, it is neither a representation of the object nor just a set of pixels, but a complex pragmatic and communicative object, a kind of an "image act" (by analogy with a speech act).

Following IIIF's terminology, a collection is made up of manifests. So, what the digital collections actually store is neither artifacts, nor even their digital representations, but rather these synthetic digital objects that bear no resemblance or similarity to the initial physical things stored in physical archives. The artificial or fabricated nature of the object in the digital archive needs to be identified not in order to criticize it and glorify the former auratic experience anchored in the materiality and authenticity of the thing. Quite simply, we need to be clear about what order of "reality" do the digital objects represent, what exactly we are engaging with.

Manipulating image

Not only do IIIF objects generate "the manipulated images", but they also make them available for further manipulations: annotating, transcribing, interacting, plotting different image sequences or gluing together different pieces.

In the digital archive, the image constitutes the major zone of interaction and contact between the user and the artifact. Most collections provide the user with the ability to manipulate the image: it can be zoomed in, zoomed out, moved, rotated, clipped, downloaded or shared through social media (thereby removing it from the collection and making one's own). Some collections[115] even feature a full-fledged environment for working with images, providing the ability to adjust different parameters (brightness, light, contrast), to mirror and flip images, to set marks on it, to cut and crop, to display transcriptions of text in parallel with the image, to annotate, etc.

Open to manipulation, the image, in this sense stands in stark contrast to the text (descriptions and metadata), which is usually inaccessible to any alteration or interaction. Offering a wide range of ways of dealing with images (indeed, "ça a été manipulé"!), the archives, however, keep under control their descriptions, built, as we have seen, on the basis of "facts" and "numbers". In this perspective, the emphasis on factuality and evidentiary value

[115] Some examples include "iManuscripta" project, "Darwin's virtual library", "Solar observations" collection.

in the textual descriptions appears as an attempt to counterbalance the manipulability and plasticity of the image. It is the text that is meant to attest to some material reality behind the "ephemeral" image, which in itself can no longer serve as a certificate of presence.

The forms of engagement with the image offered in digital collections bear no resemblance to any form of interaction with a material thing, neither *Vorhandenheit* nor *Zuhandenheit*. Take, for example, the most common feature: zooming in on an image (50% of archives). Archives make it possible to magnify an image to a degree beyond the reach of the human eye, to the point where the object is stripped of its integrity and lost from view. Instead, the tiniest fragment of a thing takes up the entire screen: like the "t" from one page of Darwin's manuscript, or the light bulb wires (Fig. 2.2.11; Fig. 2.2.12). In such a maximum approximation, the thing loses its integrity and each of its fragments gains its autonomy and independence.

Such an engagement with the thing is clearly no longer the simple (inactive) contemplation that Heidegger had in mind, but some entirely new relationship, in which the scale of the thing as well as the capacities of the human eye are no longer the point of reference.

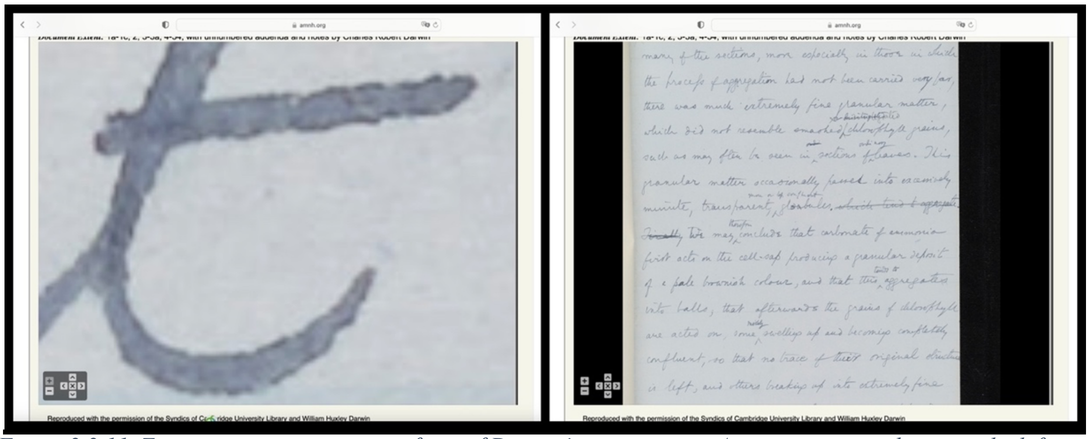


Figure 2.2.11. Zooming in on t on a page of one of Darwin's manuscripts. An entire page is shown on the left.

Figure 2.2.12. The wires of the light bulb from the Marconi collection. An entire light bulb is shown on the left. [https://hsm.ox.ac.uk/collections-online#/item/hsm-catalogue-15553]

The digital archive of science brings things into focus, provides access to things, puts things on display. Yet, paradoxically, the representation of the thing in it is emancipated, detached from the thing itself. It is no longer an emanation of the referent or proof of its presence. Instead, the digital representation acquires independence and implies entirely new and distinctive forms of engagement.

In making things visible, the digital archive offers a new (fluid and spacious) scopic regime[116] in which, on the one hand, the thing appears as a detail of an ornament or multitude of things, and on the other hand, it can be zoomed in to the point where it becomes indistinguishable. *Ornamentality* and *zooming in* both contribute to the "extension" (McLuhan, 1964) of our vision. The former enables a macro-perspective, the latter provides a micro-perspective. Both, however, are not proportionate to either the human eye or the object itself, which leads to the fragmentation and emancipation of the image from the material object itself. The digital archive, moreover, introduces a regime of extreme scalability and fluidity of vision, allowing for rapid shifts between a distant view, surveying the *multitude*, and a close-up view, examining the minutest details. These forms of visual engagements or "enhancements" substitute the physical, intimate, and tactile interaction with the archival object.

[116] The notion of the "scopic regime" dates back to the studies of film theorist Christian Metz (Metz 1975, 44). It describes "a non-natural visual order" or experience "constructed by a cultural/technological/political apparatus" (Jay 2008, 1).

2.3. On the semantics of digital connections[117]

The capacity to establish, model and represent the connections between items is widely regarded as a hallmark of the digital collections, as opposed to the order of the traditional archive. We have already observed this claim in both the structural and the media-technological approaches to digital collections. In Wolfgang Ernst's phrasing, "the new [digital] archive's task is to meaningfully link up different information nodes" (Ernst 2012, 83); in it, "aesthetics of fixed order is being replaced by permanent reconfigurability" (ibid., 99).

The digital archive no longer holds the immutable things in place, but instead is constantly re-combining, re-configuring, re-modelling the way they are ordered in relation to each other. To frame it in Bruno Latour's terms (Latour 2012), digital representations appear to be far more mobile than the original artifacts, therefore the digital archive is much more powerful in "drawing things together" than the traditional one.

If the archive both connects and disconnects, then those very connections (as well as their meaningful absences) need to be traced, mapped, and explicated. As Geoffrey Bowker and Susan Leigh Star noted in their study of classification,

> Every link in hypertext [..] reflects some judgement about two or more objects: they are the same, or alike, or functionally linked, or linked as part of an unfolding series (Bowker, Star 2000, 7).

If so, then how and on what grounds do the archives link items? What is the place of the connections among other forms of archival "categorical work"? In what ways do these ties operate and what are their implications for object representation?

These issues stand in contrast to the way in which connections are typically addressed in Library and Information Science. Although the relations have been its recurring topic and source of reflection, they are treated more in terms of knowledge organization systems and information retrieval (Green et al. 2013; Stock 2010; Boteram, Hubrich 2010; Peters, Weller 2008). LIS tackles problems such as "how to improve recall and precision of information retrieval?" or "how to generalize relations for structuring information in different knowledge organization systems (e.g., thesauri, folksonomies, controlled vocabularies)?".

Putting the focus on links and ties, I, by contrast, will consider them to be a form of giving meaning to an artifact within the archive. In other words, I will be interested in how objects contextualize each other through their linkages, how they are brought together and dissociated, and to what extent these relations contribute to their interpretation. My analysis

[117] In its condensed form, this chapter is published as Volynskaya 2024.

will begin by looking at the connections in one particular collection. I will then engage the context of the entire corpus of collections and draw about some more generalized conclusions.

2.3.1. Networking the Cavendish laboratory collection

My starting point will be a close reading of one particular archive, the Cavendish laboratory collection exhibited by the Cambridge Digital Library. Featuring photographs and correspondence from the mid-19th century to 1970s, the collection serves as a backstage of the renowned scientific institution[118]. The history of the laboratory, filled with Nobel prizes and groundbreaking discoveries, has been documented in a few monographs and is quite well known. That makes it all the more interesting to see how it is refracted and imagined in the digital collection[119].

To examine the structure of relations in the collection[120], I make use of network analysis and its metrics, so as to calculate various indices of influence, isolation, and connectivity. The nodes of the networks will be the archival objects, and the edges will be the established links between them. The technical language of network analysis allows exposing the connections and ruptures within the collection, the degree of its cohesion (*connectivity*), the importance (*centrality*) of some of the objects (*nodes*) and the exclusion (*isolation*) of other ones[121].

This dry structural analysis turns into *interpretation* once one asks how the influence is distributed among the nodes and what are the implications of this distribution. As I will show, the very structure of the collection is signifying, telling, indicative of how the institution's past is being imagined, marked out, and expressed.

---

118 The bulk of the collection consists of the photographs assembled in the 1970s by the laboratory photographer Keith Papworth. In the 2010s, several hundred photographs were selected from his catalogs to form a digital collection. The Cavendish laboratory collection/ Cambridge Digital Library [https://cudl.lib.cam.ac.uk/collections/cavendish]. On the history of collecting and exhibiting at the Cavendish lab, and the relationship of the lab's physicists to their own past, see Jardine 2019.

119 Another reason for choosing the Cavendish collection is rather technical: the collection has an ordinal numbering of artifacts (from P1 to P 2058), which allowed me to automate the data retrieval process.

120 To give a formal structural description, the collection does not make use of ontologies or controlled vocabularies. The metadata system – in particular, the subject categories and relations between objects – was developed manually in spreadsheets and subsequently converted to TEI elements. The records in the collection are described by means of a fairly common metadata grid, including physical location, place of origin, dates, creators, materials, format, and extent. In addition, one of the metadata fields used is "Note(s)", which specifies the so-called "associated images" for a particular artifact. The elaborated vocabulary of subject metadata as well as the "associated pictures" form the basis of this analysis.

121 Network analysis represents but one form of quantitative analysis of digital collections, along with different forms of visualizations and various types of statistics. For examples of quantitative exploratory analyses of specific collections, see MacDonald 2023; Gauven et al. 2017; Kräutli 2016.

In this analysis, I will be drawing on two networks[122]. The "subject network" (Fig. 2.3.1) shows how the artifacts of the collection are interconnected through *subject categories*. The "object network" reflects the direct links[123] between archival objects as they are mapped in the archive.

Subject network[124]

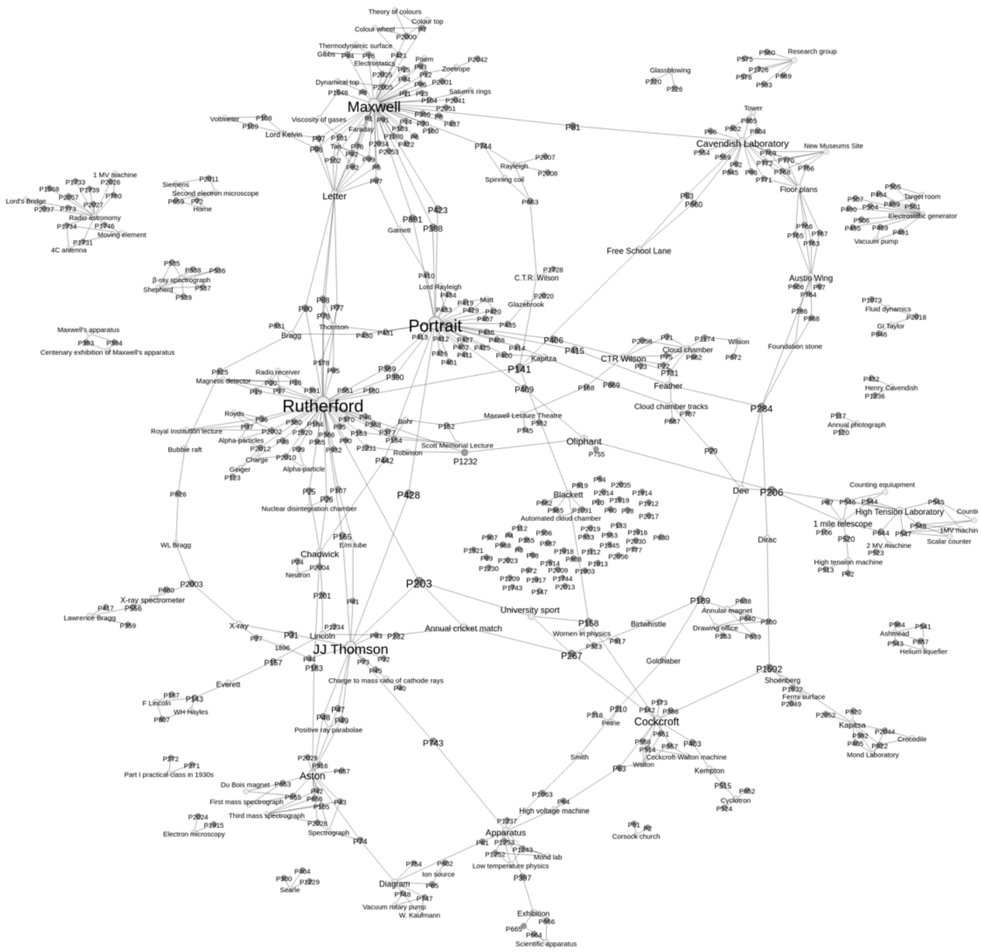


Figure 2.3.1. The Subject network. The size of a node reflects its betweenness centrality.

The subject network displays the relations between the subject categories and the objects they aggregate. This network thus exposes the mediating[125] function of the subject

[122] The networks were built in the Gephi software based on data harvested from the collection website. Both cases used the Force Atlas layout algorithm. On the network visualizations and below, I indicate the serial number of the artifacts starting with "P". All the items can be accessed via https://cudl.lib.cam.ac.uk/view/PH-CAVENDISH-P-00000, where the ending zeros are to be replaced by the corresponding number.

[123] In the digital archive, the connections between objects are identified in the metadata field "Note(s): Associated images". I followed those in constructing this network.

[124] For the data and statics of the Subject network cf. Table 4. [https://zenodo.org/records/10548106]

[125] Only subjects that link at least two artifacts are included in the network.

classification of the collection: the ways in which archival markup bridges and binds together disparate archival items and makes sense of them.

In terms of the overall structure, the subject network is made up of one big interconnected hub, that consolidates most of the nodes, along with several separate clusters grouped together by one or more subjects, and isolated standalone nodes without any ties. Statistically, more than two-thirds of archival artifacts are defined by one or two relationships, about 15% are described by three connections, and about 10% have no ties at all. More often than not, hence, the object in the collection is described by only one or two categories, which can be interpreted as a tendency toward stabilization of artifact's meaning and interpretation.

In network theory, there are several ways to measure the importance of individual nodes within a network. I will rely on the so-called betweenness centrality, which shows how important a particular node is for the cohesion of the network. The top ten list of the nodes with the highest betweenness centrality includes as many as seven proper names: *Rutherford*, *Maxwell*, *Thompson*, *Cockcroft*, *Aston*, *Oliphant*, *Dee* (the three remaining entries being *portraits*, *letters*, *Cavendish lab*). The network thus perfectly captures the lens chosen to represent the history of the Cavendish Laboratory. It is framed as a personalistic history of the "great physicists". A proper name – as opposed to an instrument, a discovery, an institution – serves as the main reference point of that history. It is the name that ties together machines, persons, artifacts, places, and objects of study. For instance, the *CTR Wilson* category links the portraits of the scientist (P406) and his colleagues (P183, P201), the machines he built (P22, P75, 1174), the particle tracks he observed (P731), and his small hut where he carried out his experiments (P1728).

Using big names for classification is quite a straightforward strategy. On the one hand, it serves the function of orientation and differentiation enabling the user to navigate through different periods and a myriad of machines, faces, and artifacts. On the other hand, the association with great scientists is also a form of legitimization and attaching value to the artifacts of the collection. The name confers aura to the object: in this logic, the magnetic detector from Rutherford's radio receiver is valuable and worthy of representation precisely because it belonged to Rutherford. In this way, the collection reflects and pursues a personalized history of science in which the names of prominent scientists serve as entry points into the scientific process.

Representing the lab's history through names inevitably obscures other facets of the story: that of instruments, of institutions, of science-in-the-making. But perhaps most importantly, by highlighting the specific names and presenting the laboratory as a collection of distinguished individuals, this strategy obscures, relegates to the background the *collective* history of the laboratory. It renders untraceable the very communal and collaborative endeavor: even the most “generic” of the statistically significant categories, *the Cavendish lab*, is used to describe building plans rather than, for example, collective photographs of the laboratory’s personnel.

Another curious effect of the personalized markup is that the most influential objects in the archive are those photographs that depict several “influential names” together. In terms of structure, the most important, backbone object of the archive turns out to be the photograph (Fig. 2.3.2) of Rutherford and Thomson at a cricket match (its exceptionally high betweenness centrality is explained by the fact that it brings together three statistically important categories at once: *Rutherford*, *Thomson* and *cricket match*). You cannot argue with statistics, yet this artifact hardly qualifies as a key to the history of the Cavendish laboratory.

Figure 2.3.2. P203: Rutherford and JJ Thomson at a cricket match

Object network[126]

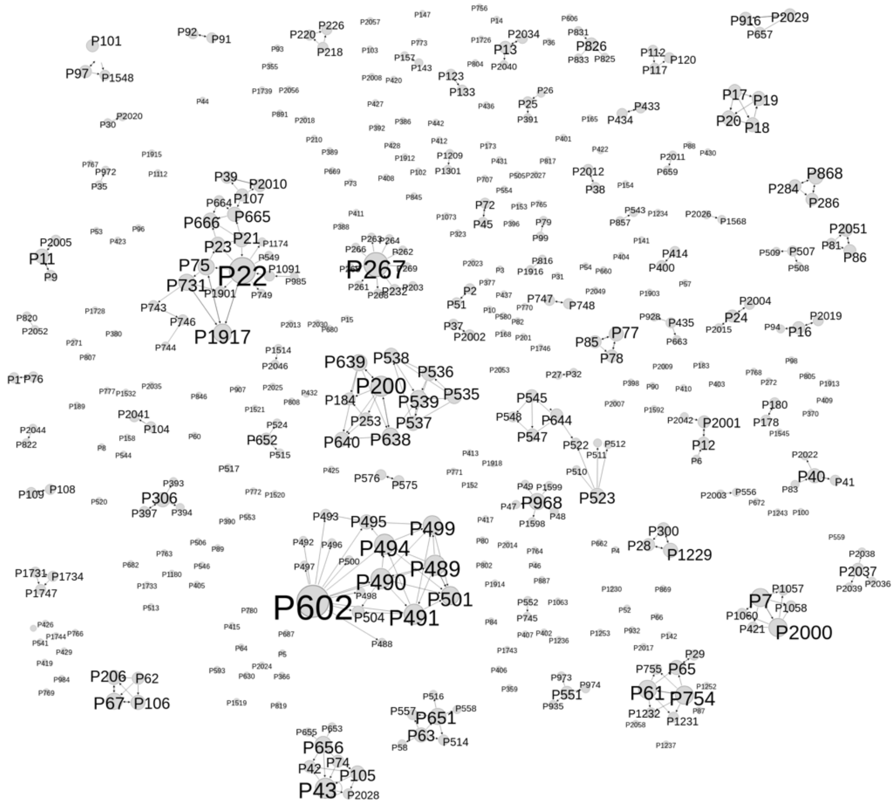


Figure 2.3.3. The object network. The size of the nodes reflects the betweenness centrality.

Let me now turn to the second network (Fig. 2.3.3), representing the horizontal or direct connections between objects. Whereas the subject network does link most of the collection, here the ties between objects do not even form a central interconnected cluster. This network looks more like a bunch of unrelated nodes with a few interconnected regions. Statistically, almost half of the network nodes (44%) are completely isolated, i.e. have no connections with other objects; another 21% have connections with one object and 16% with two objects. Consequently, the network does not represent a single coherent whole, but rather a multitude of disparate fragments, which sometimes form small regions of meanings and relationships.

The logic of connecting nodes in this network has nothing to do with the personalistic order we observed in the subject network. The history of big names here is substituted with micro-connections between devices. The archive connects either the photographs of the same apparatus (e.g., multiple images of Maxwell's color wheel, including one with Maxwell, or multiple images of the Cockcroft-Walton accelerator, including one with Walton), or various parts of one apparatus (e.g., Van de Graaff ion source, its base, its vacuum pump gauge, its

[126] For the data and statics of the Subject network cf. Table 5. [https://zenodo.org/records/10548106]

magnet, its target room, its model and project (Fig. 2.3.4.)), or apparatuses located in close proximity to each other, such as in the same room.

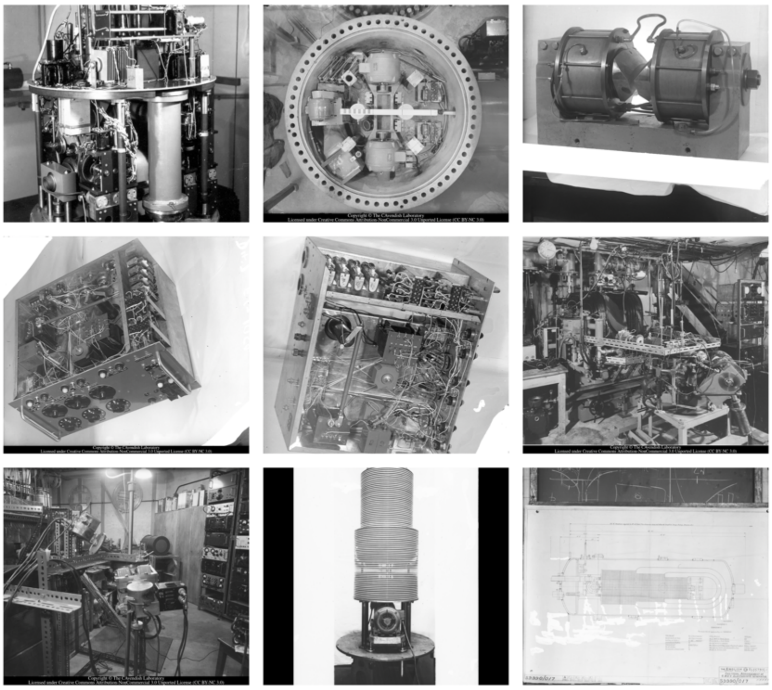


Figure 2.3.4. An example of connected artifacts in the Cavendish collection. Van de Graaff ion source and its parts.

In this choice of connections, two circumstances are of interest. The first is that none of the connections trace the persona. Only the instruments are related and *relatable*, leaving the persons (if at all in the picture) to serve merely as a background for the devices next to them. The second observation is just how static and rigid the established relations are. They only register some adjacency, the proximity of two instruments or their parts. In neither case does the connection allow for an amplification, unfolding, development, tracking changes or tracing continuities (e.g., between different machines).

Better yet, this is evident in the famous photograph "Talk Softly Please," (Fig. 2.3.5) one of the symbols of the laboratory that appears in any book about Rutherford, the Cavendish lab and in many books about history of the nuclear physics. Taken in 1932 by C. E. Wynn-Williams, the photograph shows Ernest Rutherford and J. A. Ratcliffe talking in the so-called drawing engineering office. A *Talk Softly Please* sign is lit above them, and a machine for detecting and counting particles, the so-called Wynn-Williams-Ward amplifier, stands on a cart in front of them.

Figure 2.3.5. “Talk Softly Please” (in the middle) and its related artifacts.

In the network, “Talk softly please” is linked directly to the three objects: photographs of the annular magnet (P639) and of the drawing office (P200, P253). Through these three images it is also related to the other two photographs of the same magnet, as well as to images of Shepherd’s beta-ray spectrograph. The connection therefore is predicated on two grounds: either the setting (the drawing room) or the magnet (a small fragment of which, if to look closely, can be seen in the bottom right corner of the “Talk Softly”). No more ties are made, and no more contexts are traced.

Mapping the missing connections could be a worthwhile exercise in history-of-science analysis. Here I will only point to a few of the most glaring omissions. To start with the obvious: none of the links associates “Talk Softly please” with neither Rutherford, nor his interlocutor. The magnet that binds several items together was in fact used to study alpha particles by the four physicists (Rutherford et al. 1933) each of whom had something to do with the photograph: Rutherford, Wynn-Williams, B.V. Bowden, who designed the “Talk Softly Please” sign, and W. B. Lewis, who wrote a detailed essay (1984) based on the photograph. Next, the Wynn-Williams and Ward amplifier invites one to trace both the history of its invention and modification, and the history of its use, important for counting practices in the nuclear physics. Even following the same “adjacency” logic, one could establish many more ties: e.g., connect the photo with the images of James Chadwick’s laboratory, to where the open door leads, or with the researchers who worked in this very notorious drawing office, further with the machines they worked on, and their research. Finally, there is also the context of the photograph itself, the history of its reproduction and citation.

These overlooked and neglected contexts had to be outlined so as to point to all those (numerous) perspectives and ways of making sense of the item that could have been applied in the digital archive. Instead, the archive contextualizes such iconic object as "Talk Softly Please" by appealing to spatial and object proximity, by means of a location and a piece of magnet.

To conclude the commentary on the Cavendish collection networks, I shall address those artifacts that remain isolated, i.e., have no ties in either network. There is a total of 33 such objects, representing 7.5% of the total number of archival items.

First, one can note a tendency that is widely discussed in the information sciences: the excessive granularity of categories[127]. So, for example, the portrait of J. J. Thomson (P1230) is detached from the large cluster associated with Thomson simply by the fact that among the subjects there are four variants of his name spelling. The detail of Cockcroft and Walton's 1930 accelerator (P1521) is isolated as there are at least four categories describing their experiments and equipment (Cockcroft and Walton experiment, Cockcroft and Walton's 1930 accelerator, Cockcroft-Walton machine, Cockcroft-Walton machine 1930, Cockcroft-Walton machine 1932). Maxwell's poem (P3) appears to be separated only because it is assigned to a self-standing category "Maxwell poem" which has nothing to do with the "Maxwell" category.

Second, among the isolated artifacts there are a number of technical devices or their components (P1519, 1743, 1903, 1913, 2009, 2013, 2014, 2017, 2023, 2030, 2035). These images could well have been related to others in the logic of adjacency that we discussed above, but no coherent components were identified for them.

Third and the most interesting case to address comes from a set of artifacts depicting the collective and daily practices of the laboratory: photographs of assistants, practical classes, minutes of meetings (P4, 89, 553, 808, 907, 1112, 1545, 1912). These very artifacts would seem to be at the heart of the lab's network as they depict *science-in-the-making,* routine and commonplace practices, bringing together people and machines. Yet the images of daily

127 This case is well described in information science as a tradeoff between economy and informativeness: "Thus, economy and informativeness trade off against each other. If categories are very general, there will be relatively few categories (increasing economy), but there will be few characteristics that one can assume different members of a category share (decreasing informativeness) and few occasions on which members of the category can be treated as identical. If categories are very specific, there will be relatively many categories (decreasing economy), but there will be many characteristics that one can assume different members of a category share (increasing informativeness) and many occasions on which members can be treated as identical" (Komatsu 1992, 501).

laboratory research and teaching ironically find themselves isolated at the periphery of the collection as the two selected linking frameworks fail to contextualize them.

2.3.2. Between subject and object connections

I constructed the two networks following the logic of the archive: the subject metadata and the direct object links seemed to be the major forms of creating (semantic) order linking and detaching objects. That said, these two ways of semantizing objects also echo the two facets of order as defined by Michel Foucault:

> Order is, at one and the same time, that which is given in things as their inner law, the hidden network that determines the way they confront one another, and also that which has no existence except in the grid created by a glance, an examination, a language… (Foucault 1994 [1966], xx)

Subject relations are produced by language, through words, by assigning objects to categories, while direct object relations describe a thing through its confrontation, connection, association with other things. Subjects fix the aboutness of things, while object relations only grasp some (associative) connection between two artifacts, contextualizing them through each other.

In general, subjects represent a much more common way of semantizing an archive than direct links between objects: so, out of a total of 118 collections in the corpus, 95 collections (80%) are structured through subject relations, with only 36 (30%) establishing ties between items[128] (in addition, as we have seen in the Cavendish example, these ties do not always add much meaning to the objects they describe).

Yet establishing "related objects" seem to be a special power of the digital collections, for they are able to contextualize an object beyond language, through its relationship with other objects. Let me at this point go back to our initial light bulb example, as tangible things better illustrate my argument than photographs from the Cavendish collection. Consider again the light bulb from the Science Museum Collection and its "medical context" (Fig. 2.3.6).

[128] For the data, see Table 6. Subject/Object relations. [https://zenodo.org/records/10548106]

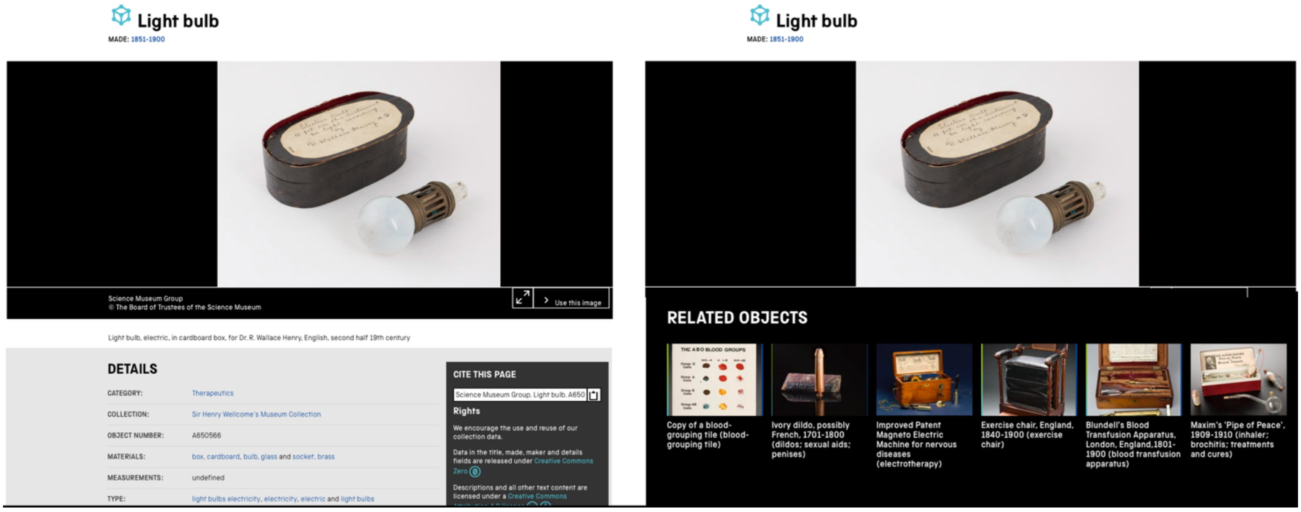


Figure 2.3.6. Two ways of describing a light bulb in the Science Museum Group collection.

Neither the formal description of the bulb, nor the subject headings, nor even its photograph give such an insight into the being of a thing, as does a selection of the related objects. The fact that the bulb is assigned to the "Therapeutics" category, is made of brass, glass, etc., belongs to the "electricity" type, and even the fact that it belonged to Dr. R. Wallace Henry and was used in the second half of the 19th century is less of a clue than *seeing* it amidst the "related" objects. All of this metadata captures the information about the bulb through an elaborated categorical language, but do not associate it with its natural context of other things. It is its juxtaposition with the "hypodermic syringe," "magneto electric machine for nervous diseases," "ivory dildo," "Blundell's blood transfusion apparatus" and other devices that marks the horizon of its use (and meaning) and gives room for the user's historical imagination. Even though some connections are quite controversial[129], it is nevertheless the context of things that allows us to imagine the light bulb in use, next to other medical instruments. In its encounter with other things, the thing becomes part of some material order (like those things we encounter in everyday life) rather than a self-sufficient and self-valuable monad, or a carrier of information, as it is often portrayed by the archive.

Another issue is the very nature of the relationship between objects, as established within the archive. In the case of the Cavendish lab, as we have seen, it is made solely on the grounds of their adjacency and material/spatial proximity. In general, this logic applies to most of the collections in the corpus.

---

[129] As far as I can discern, the sampling of these items was done automatically, via metadata. I will comment on this strategy below.

In establishing object relations, the archives typically seek to identify similarities, rather than to put things into a dialogue. The archive is more likely to bring together the two images of the same instrument than the two instruments used together as part of the same experimental set-up. In the collections, this (semantic) practice is referred to as similarity or relatedness: users are prompted to inspect "similar" or "related" objects to the object being viewed. Oxford's History of Science Museum defines "similar objects" as "some items … that share some of the same features as the object … [on view]. They may have the same name, a similar description or are from the same place". In the collections of the History of Science Museum, as well as many other authoritative collections (in 80% of those collections in my corpus that make direct connections between objects), including those of the Library of Congress, this similarity is interpreted and calculated algorithmically based on metadata. In practice, it means that, for instance, Marconi bulbs from Oxford history of science are doomed to remain neighbors exclusively with the typologically similar Marconi bulbs. For those things that have no counterpart, the logic will be more convoluted: the already familiar to us Freud's Greek statue from the Library of Congress collections stays side by side with his clock, simply because both artifacts belong to the same part of the collection called "Artifacts and paintings" (Fig. 2.3.7).

Figure 2.3.7. Pocket watch, undated; Small Greek statue, circa 6th century BC. Sigmund Freud Papers.

The archive hence seeks sameness instead of differentiation (differ*a*nce). The meaning is given to a thing through embedding it into the networks of similar objects. One problem with this approach is that since the boundaries of this similarity are undefined and controlled by algorithms, the result is often either a multiplication of the exact same things, or, conversely, an artificial linking of items that are too remote from each other. Another problem relates to the very attitude toward an archival thing, which stands behind such an approach. The archival artifact in this logic is treated mechanistically as motionless and inert, replaceable, as a commodity that can be substituted with an alternative product. This approach fails to recognize the agency of a thing: its involvement in different contexts, its participation in social relations, its influence on other agents. While Science and Technology studies engage with the agency of things and closely observe how scientific objects come into interaction with each other, forming part of set-ups and assemblages, the digital archive arranges and sorts things by their similarity and proximity.

A close observation of how the archive draws links and ties makes apparent the implicit semanticity of the archival order. As we have seen with the Cavendish Collection, the order of the archive can be mapped, read and interpreted as a form of representation or imagining of the past. It determines what is visible and what is hidden, builds implicit hierarchies, sets connections and discontinuities. As claimed by the "communicative argument", the path of the user through the archive is much less predetermined than in the narrative. Yet the archive also establishes possible trajectories, offers certain paths, develops transitions from one object to another.

As we have seen, even a simple light bulb or a portrait of scientists can be signified in many different ways. This signification zone appears to be especially sensitive where special knowledge is needed to interpret an object, such as in the case of scientific objects. To preserve such artifacts means not only to keep them intact, but also to frame the horizons of their meaning, uses, relationships and interactions with other objects. Making connections, in this sense, appears as a distinctive power of the digital archive to (re)model and to re-imagine the past. The realization of this power, as we observed, is still far from being fruitful. Archives do not take full advantage of their semantic capabilities, offering rigid adjacent or hierarchical relations.

2.4. Names and things: strategies of archival classifications

> In its remote pages it is written that the animals are divided into: (a) belonging to the Emperor, (b) embalmed, (c) tame, (d) suckling pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera…
>
> Jorge Louis Borges,
> "The Analytical Language of John Wilkins"

> *Smiling, Smoking, Eyeglasses, Radio telescopes, Waterman, Alan Tower, 1892-1967, Bainbridge, Kenneth T., 1904-1996, Blackboards, Spectrographs, Suits (Clothing), Telescopes, Bridgman, P. W., 1882-1961, Mustaches, Particle accelerators, Spouses, Tobacco pipes, Solvay Conference on Chemistry*
>
> Subject metadata of the Bainbridge collection,
> American Institute of Physics (a fragment)

The famous quote from the "Chinese encyclopedia" with which both Foucault's laughter and "Words and Things" start, perfectly exemplifies the focus of this chapter, namely the languages and orders of archival classification.

Lists of words as a way of organizing, expressing and producing knowledge have received a lot of attention in the humanities. They have been articulated as one of the earliest genres of writing (Goody 1977), as a literary trope (Eco 2009), as a way of ordering human interactions (Bowker, Star 2000), and even as an ontological category (Bogost 2012).

In Library and Information Science, the technologies of indexing and classification serve not only as an object of reflection but also as a site of practice. As defined by Wolfgang Stock, "indexing means the practical work of representing the thematized single objects in a documentary reference unit on the surrogate, i.e. the documentary unit, via concepts" (Stock 2015, 527).

According to information scientist F. W. Lancaster, the indexing process is guided by two concerns: 1) the aboutness of the item and 2) the relevance to the user (Lancaster 2003, 9). Indexes (or subject categories/keywords) thus constitute both semantic and pragmatic markers. On the one hand, they are intended to describe the object, being "the unit of meaning", identifying the most important point, reflecting what the item "is really about, avoiding trivia and insignificant detail" (Scott, Tribble 2006, 55). On the other hand, indexes and subject classifications are "message-oriented elements" (Bondi, Scott 2010, 7) meant to guide the user through the digital archive. They are designed to function as an area of communication with the user, to both anticipate the user's query and respond to it.

The primary question this chapter addresses is how are these semantic-pragmatic markers organized within the collections? What strategies and classification frameworks do the collections of science tend to adapt? And how do they organize the user's reading of or navigating through the archive?

Lists of categories are not designed to be read sequentially and thoroughly from first to last word. They suggest rather a quick, cursory, surface glance through the contents until the eye catches the category of interest. What I am about to do in this chapter is to make them a full-value object of reading, to read them from the first to the last word in order to recognize the classification strategies being used in the collections.

The technical part of the study is quite simple. First, I gathered[130] all the subjects from each of the collections. I then sorted them into three files: categories in (1) personal collections, (2) institutional collections, and (3) collections of scientific instruments[131]. This gave me three vocabularies (or rather "bags of words") representing different types of collections, that could be read and interpreted. I then applied to these vocabularies a sort of manual topic modelling, classifying the individual words into categories I had established (such as "proper names", "disciplines", "theories", "objects of study")[132]. With such categories, one can trace the frequencies of particular discourses or indexing strategies (rather than individual words) and their distribution across the collections. The statistics will serve as an entry point into the discussion of each type of collections, allowing me to capture a pattern to be further interpreted and complemented by specific cases[133].

2.4.1. Personal collections: *Paul Henneberg*, *mustaches* and *innovation*

Mario Biagioli once suggested to consider the name of a scientist as a *document* essential for "the workings of the economy of science" (Biagioli 2006, 127). Tracing naming strategies in scientific publications, Biagioli showed that scientific claims and scientific credits appeared to be attached to author's names.

---

[130] Only collections containing subject categories were considered – there are 95 such collections in the corpus. The method of collecting categories depended on the collection interface. In some collections, they were collected manually by simple copying, while in some collections they were gathered with the help of code.

[131] The division into collection types corresponds to the collection description in Table 1. I excluded thematic collections from the analysis since their categories are too dependent on the collection's subject matter.

[132] The data and the "topic modelling" are presented in Table 7. Subject classifications. [https://zenodo.org/records/10548106]

[133] Some of these cases were discussed in Volynskaya 2020.

Likewise in personal collections, the proper name appears to be the most common form of documentation and categorization. Personal names make up more than a third (37%) of the categories for classifying scientific residues. In comparison, names of institutions account for 9% of the categories, disciplines for 4%, descriptions of personal and work life together for 3%, theories and epistemic objects for less than 1%. We can thus conclude that in general, (personal) scientific residues tend to be classified through the names of people.

This strategy of using proper names as classifiers reenacts what Michel Foucault described as the author's function. As Foucault explained,

> [The author's name] is functional in that it serves as a means of classification. A name can group together a number of text and thus differentiate them from others… We can conclude that, unlike a proper name, which moves from the interior of a discourse to the real person outside who produced it, the name of the author remains at the contours of texts – separating one from the other (Foucault 1977, 124).

Converted into a thematic category or a label, a proper name turns into a function for classifying archival objects. It does not so much refer to a specific person outside the archive, but rather divides or brings together archival artifacts.

A record-breaker for the frequency of proper names is the Albert Einstein Collection, exhibited by The Shelby White and Leon Levy Archives Center (the Institute for Advanced Study). It is a modest collection of about three hundred photographs "of and related to Albert Einstein" indexed through 183 categories. 89% of these categories represent the proper names of individuals (combined with the names of organizations, the percentage rises to 97%!). As one can readily guess, the category grouping the greatest number of items is *Albert Einstein*.

This means that, paradoxically, on the level of semantic metadata, Einstein's personal collection does not actually provide any context – e.g., that of the Einstein's biography or his scientific claims. Instead, what it does is *labeling* Einstein and other persons. The user, in other words, cannot retrieve all the Einstein photographs of the American period, or all the photographs related to the theory of relativity, but is able to locate all the photographs featuring *Harry Camp Clark* or *M. S. Rosenblum*.

Structurally, this strategy results in a disproportionate distribution of subject categories. While most of the descriptors only cover one archival item, a small fraction of categories accounts for an excessive number of artifacts (such as *Albert Einstein* category).

Semantically, this strategy leads to even more striking effects. Consider, for example, the photograph "Albert Einstein arrives in San Diego, California" (Fig. 2.4.1). It is described quite thoroughly in the collection: in 1930 Einstein and his wife arrived in San Diego, where they were greeted by, among others, the then mayor of San Diego, Harry Camp Clark, and the "Director of the San Diego High School band", Paul Henneberg.

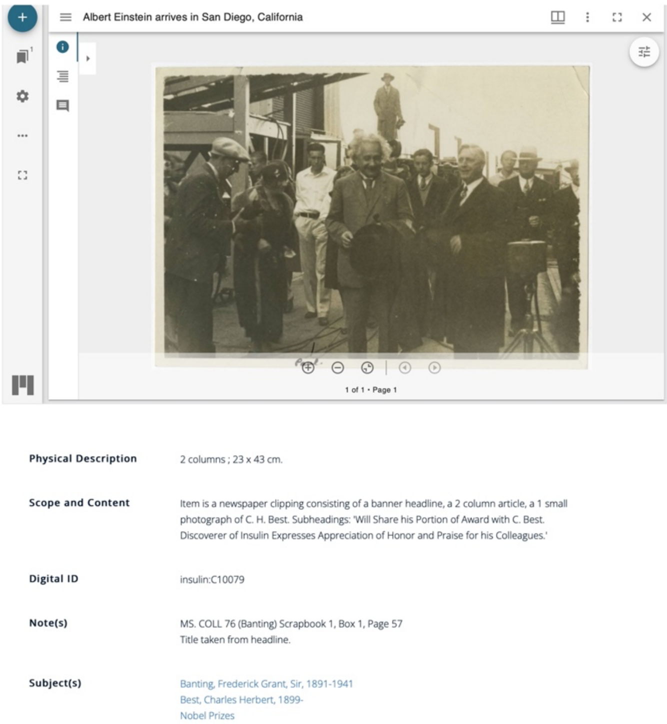


Figure 2.4.1. Albert Einstein arrives at San Diego, California, The Shelby White and Leon Levy Archives Center, [https://albert.ias.edu/entities/archivalmaterial/613b1c73-525b-450e-811c-b6c96a6a063d]

The photograph is indexed by four categories, all of which are the names of the persons depicted: *Einstein, Albert*; *Einstein, Elsa*; *Clark, Harry Camp*; *Henneberg, Paul*. Significantly, the last two categories in the archive describe only one object: this very photograph.

This example provides a good illustration of indexation being driven by the pursuit of exactness and comprehensiveness of naming. All other potential descriptions and contexts are avoided, while names are given with the greatest possible accuracy and precision. As a result, the name of a high school orchestra director who once met Einstein at the train station turns out to be the keyword – "the unit of meaning"– in Albert Einstein's personal collection.

The prevalence of *naming* over other classification strategies reflects the general tendency of archives to stabilize, fix signifier as opposed to both signified and connotation. As a label, the name pinpoints what is depicted, but hardly elucidates it or contributes anything to it.

Interestingly, not only names but also things can be used in the same capacity, i.e. for the same "naturalized" and punctual matter-of-fact characterization. At this point, let me take another example – a joint photograph of J. J. Thomson and Ernest Rutherford and the way in which it is classified in two collections: the already known to us Cavendish laboratory collection and the Bainbridge collection, exhibited by the American Institute of Physics (Fig. 2.4.2).

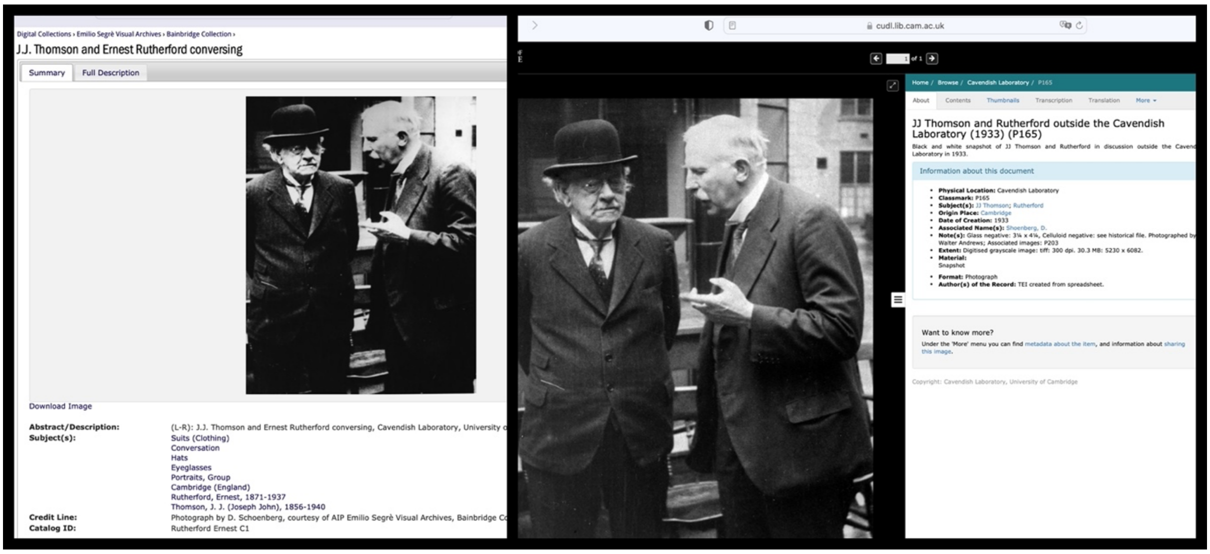


Figure 2.4.2. The photograph of J.J. Thomson and Ernest Rutherford in the Cavendish Laboratory collection (on the right) [(https://cudl.lib.cam.ac.uk/view/PH-CAVENDISH-P-00165/1)] and in the Bainbridge collection (on the left) [https://repository.aip.org/islandora/object/nbla%3A311121)]

In the Cavendish Laboratory collection, a photograph is indexed through proper names only, meaning users can find it if they are looking for Thomson, Rutherford, or both. In the collection, this photograph will only be associated with other objects related to Thompson or Rutherford. Notably, this photograph could be indexed (and *semantized*) in many different ways: as an image representing Nobel laureates in physics or the daily life of the Cavendish Laboratory in the 1930s, or any other topics. Naming, by contrast, only serves to accurately document and secure what is depicted.

Now, consider the way the same photograph is indexed in the Bainbridge collection – another modest photo collection related to the American physicist Kenneth Bainbridge. In this case, the scientists in the photograph are indexed as people in hats, glasses, and suits having a conversation and labeled *Thompson* and *Rutherford*. In the collection, this photo will be adjacent not only to all other images of Thompson and Rutherford, but also to all other persons in glasses, hats, and people having any kind of conversation.

Looking at the overall logic of the categories in Bainbridge's collection, 57% are occupied by the names of people, 12% by the names of places and organizations, and the remaining third of the categories consist of *things* or actions seen in photographs. This latter third, as we have seen, include very detailed descriptions of behavior and physical appearance, including: *Full-face, smiling, smoking, eyeglasses, hats, coats, eating and drinking, suits, mustaches, tobacco pipes*. Along with those, however, it also covers the science-related objects featured in the photographs: *spectrographs, telescopes, particles accelerators, equations, classrooms, numerical solutions*. As a result of this strategy, the photo of Joseph Mattauch speaking at a conference on Nuclear Masses is categorized under the following descriptors: *eyeglasses, blackboards, classrooms, equations, numerical solutions, lectures and lecturing, Mattauch, Joseph* (Fig. 2.4.3).

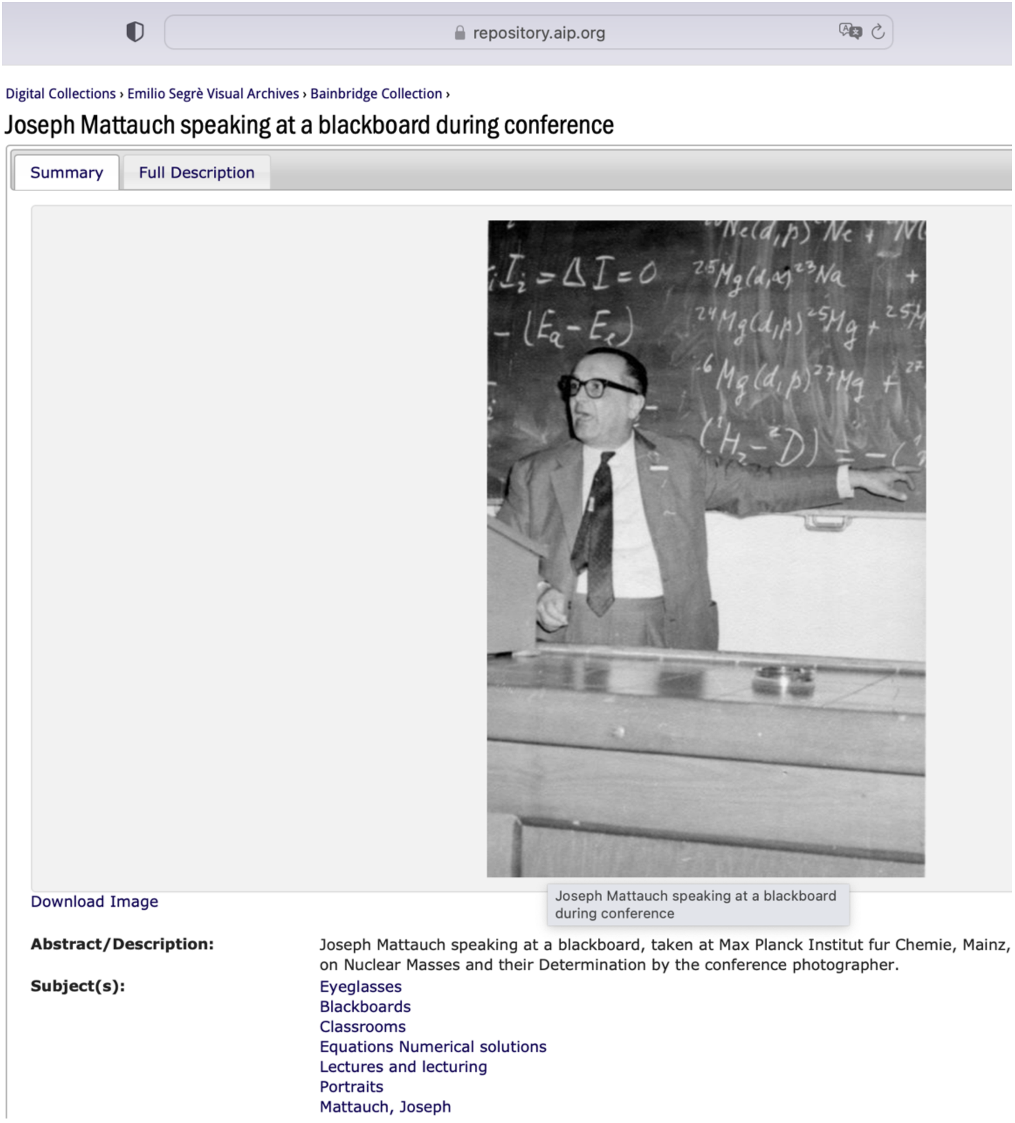


Figure 2.4.3. The photograph of Joseph Mattauch at the conference, Bainbridge collection, [https://repository.aip.org/islandora/object/nbla%3A303471]

In terms of information retrieval, this strategy brings about the possibility of querying the photograph collection based on the presence of glasses, beards, and tobacco pipes. From the perspective of knowledge organization, the resulting system is nothing inferior to the Chinese encyclopedia, featuring equations, spectrographs and accelerators alongside mustaches and hats. Should we finally assume that the classification system in the archive gives meaning to the artifacts being represented, then in the Bainbridge collection, scientists gain meaning insofar as they smoke, smile, wear glasses and suits.

In addition to names and things, concepts can also serve for the same "naturalized" and "objective" description. At this point, consider, for example, the fragments of categorical systems in the collections of two theorists and practitioners of management: Frederick Winslow Taylor and Peter Drucker.

The first is the ideologist of "scientific management" of the turn of the century, who gave the name to a whole movement in the organization of work – Taylorism. The second – "the father of modern [post-war] management" – developed the ideas of decentralization, management by objectives, knowledge work and information society.

The classifiers of the Taylor collection include:

> *labor, industrial relations, shop management, industrial management – recreation, industrial management – moral and ethical aspects, employee morale, labor, alcoholism and employment*[134]

The Drucker Collection is indexed via the following categories:

> *knowledge workers, management by objectives, decision making, leadership, performance, innovation, learning, blue-collar workers, knowledge and learning, communication, influence (psychology), globalization, motivation, compromises.*

Taken together, these two sets of subject categories provide some insight into how distinct were the management systems practiced by Taylor and Drucker. However, where the two snippets are very similar is the way in which they use the keywords.

In both cases, the classifiers are actually *concepts* removed from the theoretical/discursive systems of the two management theorists. *Labor* in Taylor's collection, or *Innovation* in that of Drucker, do not refer to the *facts* of labor or innovation, but rather to the *uses* of these terms. Thus, in the case of Taylor, *labor* refers to the letters discussing labor. In Drucker's case, *innovation* describes a variety of documents related to Drucker's book titled "Innovation and Entrepreneurship" or the presentation of an award "For non-profit innovation".

In this regard, concepts here operate just like the proper names discussed above. They perform the same function of labeling, only this time registering the occurrence of certain words. However, unlike names that are labels by nature and have no visible connotations, *innovation* or *alcoholism and employment* are politically connotated and gain meaning only within the context of the theories, systems, and discourses. Being isolated and taken out of context, *innovation* or *labor* are made into "neutral" and "objective" classifiers. As a result, the conjunction of alcoholism and employment, which takes on meaning in the context of Taylor's

[134] Hereinafter, the original spelling and punctuation of the subject categories have been retained.

project of scientific management, gains legitimacy and appears in the subject list of the Stevens Institute of Technology collection right between the *University of Illinois* and *Ray Stannard Baker* (Fig. 2.4.4).

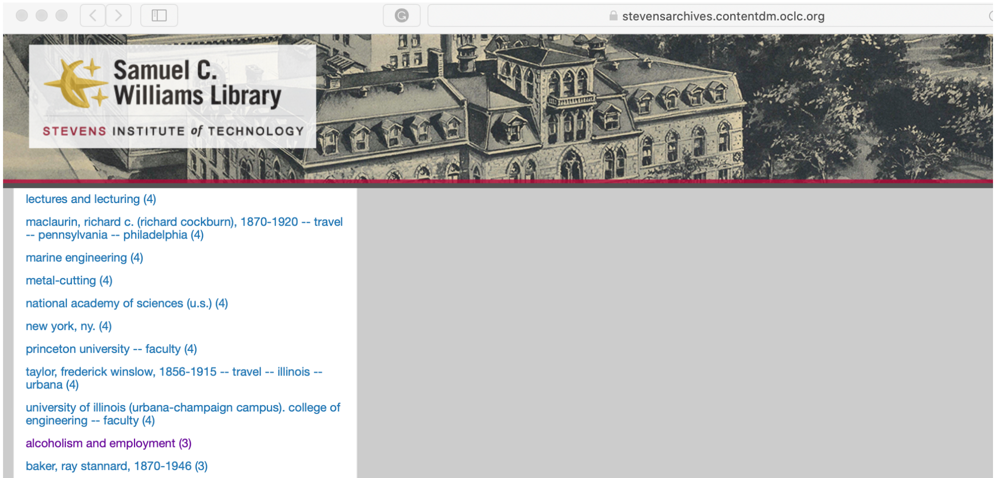


Figure 2.4.4. The Alcoholism and employment category within the subject list of the Stevens Institute of Technology collections.

2.4.2. Institutional collections: *sun*, *astrophysicists* and *women in science*

Institutional archives represent a more complex case, where different classification logics may contradict each other within even a single collection.

Let me start again with the now familiar *naming* strategy (in the case of institutional histories, the names of institutions join the names of people). Once again, it represents the most common classification strategy: proper names account for almost 40% of the categories.

In the Yerkes observatory collection, for example, nearly 90% of the subjects is made up of people and organisations. The remaining 10% consists of astronomical bodies and objects (comets, galaxies, meteoroids, etc.) as well as several categories on solar eclipse expeditions. The institutional part of the list is clearly structured: each institution is described in one category, regardless of the number of objects it refers to (e.g. *Pulkovo Observatory Buildings, Instruments, Equipment, Grounds* form a category). As far as the social and material history of the observatory is concerned, the institutional setting turns out to be the main reference point. No separate categories are defined at Yerkes for astronomers, domes, laboratories or telescopes – the only point of access to them is through a specific institution.

Interestingly, however, the institutional framework is being sidelined when it comes to astronomical phenomena. In this case, the main focus moves to the object itself: a photograph of the sun taken at the Mount Wilson Observatory comes under the category *sun* and not under the *Mount Wilson Observatory*. As far as the scientific process per se is concerned, categorisation is based on natural objects, rather than on institutions. The archive therefore maintains two separate and detached frames governed by different logics: an institutional history and a "scientific" one (Fig. 2.4.5). Both of these narratives, however, carry on the same labeling strategy as observed in the example of personal collections.

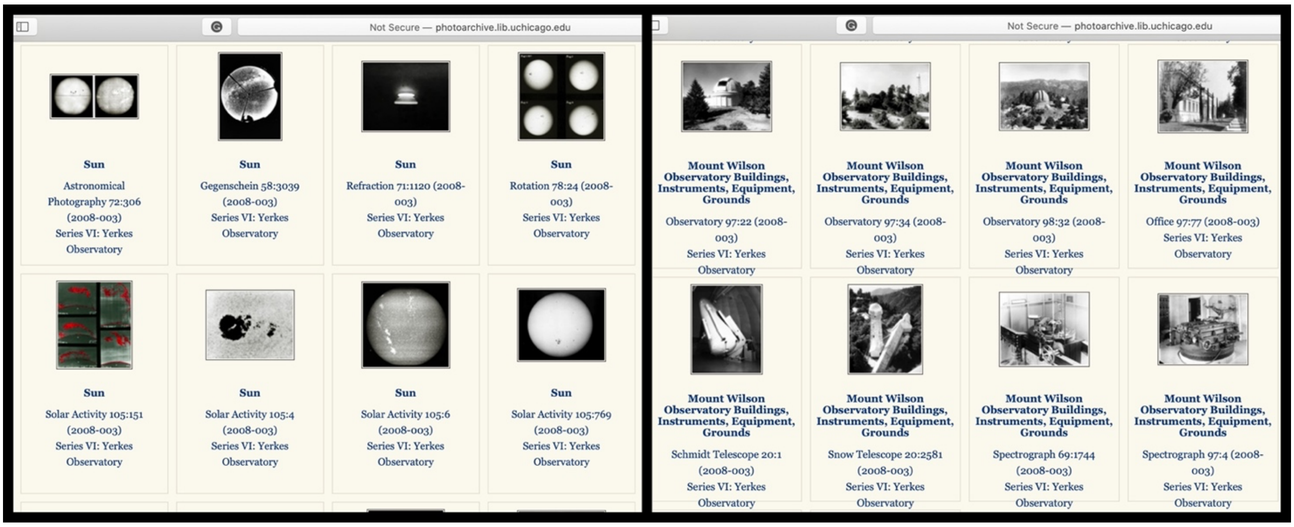


Figure 2.4.5. Photographs grouped together under the Sun category (on the left) and those grouped together under the Mount Wilson Observatory category (on the right), Yerkes observatory collection.

If we look further at the classification by "astronomical object", we realize that it does not avoid inconsistencies and collisions either. As such, this strategy clearly puts the logic of astronomers above that of historians. It stems from an interest in the object, not in the history of its representations, e.g., ways in which astronomical bodies have been captured and recorded since the late nineteenth century. The very fact of (scientific) representation fades into the background, while the emphasis is placed on the properties of the natural object, be it *solar eclipse* or *galaxy clusters*. The archive adheres to the imperative of consistency and orderliness of scientific classifications, avoiding to bring together objects as different from an astronomical point of view as the sun and the moon. However, in keeping with the logic of the scientific library, it readily draws together, for instance, photographs of different epochs.

Whereas Yerkes subject classification is more suited for astronomical propaedeutics than for the study of scientific inscriptions, the item-level description reveals the inverse logic (Fig. 2.4.6). The metadata describing a particular image of the sun, after all, is silent on the properties of the depicted celestial body, instead emphasising the fact of representation and treating the photograph as a physical object (detailing its photographer, date, physical format, etc.). It does not guide a user's gaze, nor does it allow users to interpret these images in the way astronomers could do it. The metadata does not provide any research context: what was the task of observing, what was discovered, what was known at that point, and what was not.

So, the two grounds for description – on the one hand, the categorisation by type of celestial bodies, which promises the user at least an astronomical exercise, and on the other hand, the logic of archival records, detailing the provenance and creation histories – confront and hinder each other.

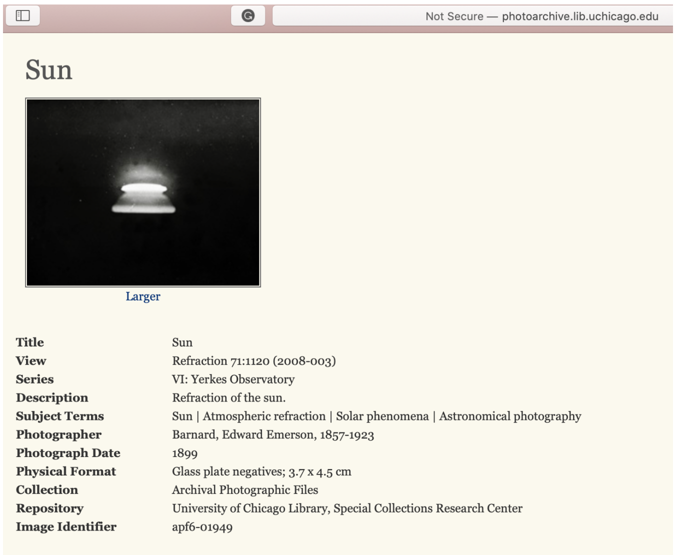


Figure 2.4.6. Example of an item-level description in the Yerkes Observatory collection.

Another and even more remarkable example of a clash of logics can be traced in classifications that rely on disciplinary divisions. In institutional and scientific instrument collections, disciplinary grids are also a rather common categorization strategy. The use of disciplinary divisions appears as an attempt to bring a scholarly order into a fragmented and rhizomatic digital archive. These classification practices seek to reduce the contingency of both the archive and the scientific process that they describe. Instead, they embody the notion of a scientific archive as a universal system of knowledge reflecting the *arbor scientiae*, where each piece of knowledge is located in a strictly designated place and is subordinate to a certain branch of science. This arborescent conception of knowledge corresponds to a particular paradigm of knowledge organization, namely hierarchical taxonomy. An extreme case of this strategy is an ultra-detailed tree structure of the INRIA collection, comprising no less than 1126 subjects, 123 of which represent disciplines.

In some cases, though, social and ethnic categories intrude into this rigid and hierarchical "scientific" classification systems. The archive of the American Institute of Physics, for example, supplements *physicists*, *astronomers*, *astrophysicists* with marital and gender categories such as *women in science*, *spouses*, *women physicists* and even *physicists' spouses*. In the Lick observatory collection, in addition to *astronomers*, *physicians*, *librarians*, *biologists* there are *African Americans*, *children with disabilities*, *older people* and many other social categories. The female categories are particularly common in the collections: there are *women* (MIT, Lick, INRIA, L'ONRSII), *mothers and daughters* (MIT), *women in physics* (Cavendish), *women college students* (MIT) and so on.

That said, no unified, general, and symmetrical grid of categories is established: having an *African American* category does not imply that there is a *European American* category, nor does having a *women* subject suggest the presence of *men* or any other gender categories. In terms of the politics of equality, such asymmetries are quite controversial. Firstly, the recognition of only a few minorities stigmatises and makes even more invisible the other, unidentified ones. Not to mention the fact that the scientific milieu has specific marginalized groups, such as the 'invisible technicians' (Shapin 1989), which none of the archives explicitly identify. Secondly, in a way, such categories imply a certain image of the majority—essentially white and masculine. Separating women or African Americans as a category is meant to uncover the presence of the marginalised groups in the history of science, yet the effect can be quite the opposite. By singling out, for instance, women and not identifying men, the archives represent women's history as a peculiar marginal case in science, as a sideline of a "big mainstream history" of knowledge, constructed implicitly, out of categories.

Apart from the equality politics, the social and communal categories bring out very clearly the polyphony of perspectives and discourses in the digital archive of science. Whereas *physicists* or *astronomers* are established in the logic of disciplinary grids, *mothers and daughters*, *older people* or *African Americans* obviously violate this order of identification. While the curious category *physicists' spouses* appears as an attempt to remain within strict disciplinary divisions, *women in science* from the same collection apparently undermines these efforts.

The social, ethnic or gender categories are often made as terse narrative formulas that participate in and respond to the debates that take place *beyond* the archive. So, for instance, the widespread inclusion of female history in scientific archives is clearly a response to the recent gender criticism and the demand to acknowledge the presence of women in science. In this multi-faceted struggle for identity, the categories do not only acquire meaning *within* the archive classifications, but also appear as a performative response to certain discussions and discourses *outside* the archive. The fragments and remnants of these discourses are now forming part of the archival subject structures.

2.4.3. Scientific instrument collections: *planimeters*, *ZDC*, *gravity*

Concluding the overview of thematic categories in collections, we shall also take a look at the forms of categorization in collections of scientific instruments. In general, the same conventional pattern is at work here. In this case, the predominant indexing strategy consists of naming instruments and disciplines: 28% of all subject categories are occupied by disciplines and 37% by instrument names. In comparison, objects of study account for 3% of all categories, research methods for 2%, and theories for less than 1%.

In most instances, therefore, collections use categories such as *astrolabes*, *compasses* or *photodiodes*, showing the very same tendency towards naming and stabilizing the signifier. This order also brings us back to the idea of similarity discussed in the previous chapter: in it, light bulbs will sit with other light bulbs, calorimeters with calorimeters, planimeters with planimeters, and hardly ever will different types of instruments meet each other in the frame of the collection.

Ernest J. Breton (1981 and 1991) criticizes this indexing strategy appealing to the fact that it fails to meet the needs of engineers and ultimately inhibits further innovations.

As Breton explains,

> What inventors (and engineers) need is a knowledge-focusing lens that selectively extracts information needed to solve their problems from available knowledge (Breton 1991, 174).

According to Breton, indexes should reflect not the names of devices, but rather their functions or properties, so that engineers can query the instruments that meet certain criteria. To give Breton's own example: a diamond saw should be indexed not as a sub-species of saws, but according to its function, that is cutting hard materials. As such, it should be grouped together with lasers, water jets and other devices performing this function (ibid., 174).

In proposing functional classification as a solution, Breton articulates the problematic of inventions (as opposed to residues) and emphasizes the viewpoint of the engineer (as opposed to that of the historian, the sociologist, the general public). In the same logic, for instance, indexing through naming instruments is equally unsuitable for presenting science to the general public, probably finding it not so easy to distinguish a calorimeter from a planimeter, and making sense of these distinctions without some additional context.

Extending Breton's point, it is worth raising the question of the pragmatics of classification and the addressee of the archive more broadly: to whom and for what can a particular indexing strategy serve? Those examples of classifications we have seen before – favoring "universal" and "objective" forms of description – are hardly addressed to any particular group or driven by a particular concern.

To conclude this chapter, I will take a look at two examples of collections that in one way or another articulate the pragmatics of categorization and its addressee. Both of these examples – more exceptions, rather than rules – provide grounds for thinking about what can we know through the archive – the question that will be taken up in the following chapters.

The first of these, the CERN collection, exemplifies an archive addressed to a specific scientific community and intended for internal rather than external use. It is one of the few examples of collections organized through the folksonomy, that is all the registered users (as a rule, CERN employees) are able to put tags, add descriptions and contribute to the classification of the collection. As a result, the collection index turns out to be immense, irregular, disordered and furthermore completely opaque to the public outside the CERN community. For example, in its classifications, CERN widely uses acronyms and abbreviations: "internal" names of accelerators (*LHC*, *SC, FCC*), detectors (*ITS, VELO, RICH, CEDAR*), experiments (*NA62, ATLAS*), specific instruments (such as *ZDC* – zero-degree calorimeters). But even those categories that, at first glance, seem to be less encrypted can be understood only in the context of CERN's activities: so, for example, the *event display* category describes exclusively the events of particle collisions (Fig. 2.4.7).

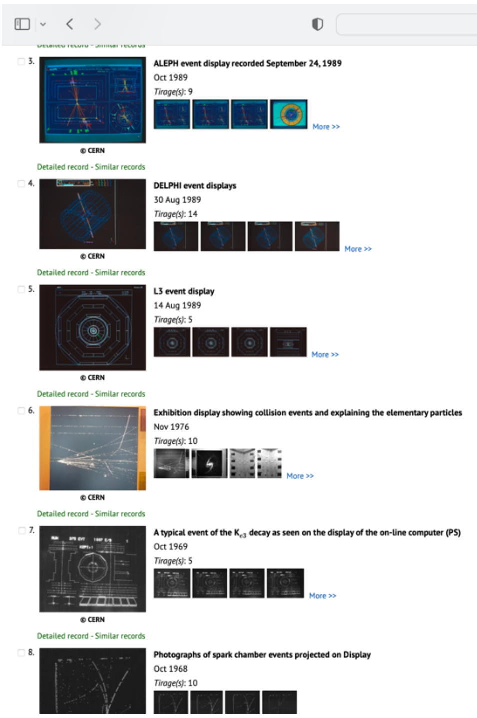


Figure 2.4.7. Examples of artifacts within a category "Event display", CERN Photolab Archives

The CERN collection is not addressed to the general audience, nor, for example, to engineers in general. It is targeted primarily at the CERN community itself, and more broadly at the professional community of nuclear physicists. The collection in this case forms part of the everyday life of the scientific institution, and its indexing appears to be simply the reflection of the community language. Made by scientists, this classification brings up the question of translation and highlights the gap between the (obscure and over-specialized) language of modern science and the public. Yet, it appears to be a form of *communication* within CERN, and will eventually articulate the community's collective memory, produced through a digital archive[135].

If the CERN collection is categorized by scientists for scientists, the second example I would like to give is that of an archive classified by historians for historians. There are several such projects in the corpus[136], but I will take a closer look at the Database Machine Drawing (DMD). It is a small collection of technical drawings from the early modern period, launched by Max Planck Institute for the History of Science in 2006.

The project founders designed a special vocabulary and conceptual apparatus for the analysis of engineering drawings, which formed the basis of the navigation system and metadata of the collection. Alongside names and titles, the DMD indexing system includes specially designed descriptors for the bibliographical source, the image and the pictured device. So, the project categorizes drawings by its intended addressee and intended purpose (ranging from *artisan commissioner* to *public*, and from *advertising* to *theoretical consideration*), by the pictorial language and graphic techniques. It also allows querying on a variety of attributes of the depicted device, including its field of employment, source of power, machine parts, gearing and drive combinations. In this way the user can locate, for example, all *engravings* made for the purpose of *advertising one's device*, or all the images of devices in the field of *water-engineering* that are powered by *gravity*.

In this case, the index system does not simply capture the aboutness of the drawings, but appears as a system for providing them meaning. According to this system, images can be placed next to each other in different logics and can be read based on different contexts. More than that, the classification of the archive, made with an eye to research, itself becomes part of the *research* practice. In the DMD case, indexing appears as more than just carrying out the inventory that *precedes* the research, but a form of inquiry, operating by links and connections. The classifications and search options proposed in the DMD project do not simply capture what

135 Such a configuration will be discussed in detail in Part IV with regard to the problematics of (digital) memory and born digital scientific residues.

136 An analogous example would be, in particular, the large-scale and decade-long Casebooks project made by a team of historians at the University of Cambridge. The project focuses on the medical records of two 16th-17th century astrologers.

the collection is about, but form a methodology that invites the technical drawings to be read in a particular way.

This part opened with the premise that archives do not tell stories. Yet, as we have just seen, the digital archive may well offer some frames and methodologies for reading its artifacts. The systems of semantic categories mark out scientists' stances, languages of instrument description, or institutional histories, even as they adhere to the most literal, strict, objective descriptors.

Yet, this semantic order remains implicit. The poetics of alphabetical lists, number-based sorting, formalised descriptors convey the illusion that the archive simply presents things as they are. However, it is the structure of subjects and the choice of descriptors that constitute the most sensitive areas in terms of politics. Through the metadata, fragments of various discourses penetrate into the archives and being deprived of context, take the form of a statement of fact.

Any category grid puts light on certain aspects and obscures the others. In this regard, the most significant absence in the categorisation of the digital archives that I have examined is the lack of concepts from the history, sociology or philosophy of science. The collections, as a rule, do not employ the vocabularies of the humanities and do not elaborate a language that would suit the residues of science, render them readable, or frame their meaning.

Conclusion

Digital collections are often defined through the notion of openness promising “open access”, “open information” and “open knowledge”. And, indeed, as we have seen, this idea of openness and transparency is embedded in the very visual *dispositif* of the digital collection as media. The digital archive unveils what has been hidden in dusty closed vaults, cabinets, cases, and brings out into the public eye what has been offstage and inaccessible to view. Semantic markup and visual forms of collection organization – the fruit of ongoing reflection on optimizing systems of knowledge organization and information retrieval – are designed to keep objects most visible and easily findable.

That said, we can hardly speak of the digital archive opening up objects in the way Jacques Derrida (1972) or Umberto Eco spoke of the “open text”: open to interpretation and decoding. In this sense, the residues of science exhibited in the digital collections remain, by contrast, tightly *closed* and inaccessible to interpretation. We have seen some evidence of this at different levels. Archival descriptions gravitate toward emphasizing the authenticity of things rather than putting them in context. The links between objects are mainly established on the basis of similarity and do not account for any other form of relations. The languages of the archive tend to naturalize and stabilize what is depicted (the signifier) rather than contextualize it.

In the end, what we observe is a great gap between the (epistemological) attitudes of the humanities and the archive; between the way the scientific artifacts are interpreted in the Science studies and the way they are represented in digital collections. While the history and sociology of science “draw things together” focusing on assemblages, networks and mobilities, the archive rather frames artifacts as things in themselves.

That said, we are also facing a clash between the order of the traditional archive and the principles, or as Wolfgang Ernst would say, *archē* of the digital archive as media. Many of the practices and notions of the traditional archive, which lie at the very heart of the archival profession, lose their relevance and power when transposed to digital collections. On the other hand, those (semantic) possibilities offered by the digital infrastructure, as opposed to the traditional archive, do not come to fruition either.

As a new infrastructure for historical research, the digital archive calls for rethinking some core concepts of the archival and information science. This even applies to such a crucial notion as *preservation*. If the archive no longer keeps tangible things but rather displays mobile and manipulable digital entities, then what exactly are we preserving and what does it even mean *to preserve*? The OAIS model, as we have seen, answers this question, but does not elucidate much else. For example, how the archival description should function in such a case? To what extent should it be tied to the material thing and be guided by traditional notions of its authenticity? Such questions need to be asked and reflected upon if we are to gain a sense of the epistemological possibilities and limits of digital collections.

The next part seeks to offer an example of such reflection. It will argue that the archival attitude towards objects should be changed from representing the thing as it is, emphasizing its uniqueness, singularity and fragility, to the thing as a force of agency that accumulates histories, models social relations and mediates our experiences. For, as Bruno Latour reminds us, things "are much more interesting, variegated, uncertain, complicated, far reaching, heterogeneous, risky, historical, local, material, networky than the pathetic version offered… by philosophers" (Latour 2005, 19-21). And by archives, I should add.

Part III. Interrogating scientific artifacts through semantic modelling

3.1. Let the stones speak: interpreting scientific residues in the archive

As witnessed throughout Part II, the digital archive makes scientific residues *visible*, yet not *legible/readable*. The archive opens things to the gaze, allows one to search for and retrieve artifacts, but does not open them for interpretation. In this part, I address the logical follow-up questions stemming from this premise: How to make the archival residues not only visible, but also readable and interpretable (for the history of science)? How to make things talk and give them meaning? And what exactly can they tell us?

This part advocates and puts forward the vision of the archive as a *librarium* of science (Ihde 1999), which preserves scientific artifacts through their contextualization and historicization. In other words, I suggest that the archive should be actively involved in interpretation of the items it preserves. While this point may seem quite provocative, it is in fact quite consistent even with some of the tenets of the traditional archive: namely, the idea that the archive contextualizes its records in a certain way (e.g., through maintaining the "original order" or adhering to the principle of provenance). The *dispositif* of the traditional archive can indeed be viewed as a mode of interpretation of the items it holds. The only (yet crucial) difference is that within the traditional paradigm these archival principles or "forms of interpretation" are seen as the sole valid and *objective* methods.

The view of "interpretative archive", certainly, aligns much more closely with conceptions of the digital archive, particularly as described in the OAIS standard (cf. Chapter 1.2). If the archive's role is to maintain and preserve the intelligibility of objects, then it has to define a certain interpretive horizon that it seeks to guard. In this logic, the preservation of objects of knowledge, such as instruments, laboratory notebooks, or research inscriptions, should encompass the consideration of what it means to maintain the intelligibility of these artifacts. In all these cases, what is of value is not only the material objects themselves, but also their history of usage, engagement in the production of knowledge and in certain scientific culture. Therefore, as I argue, preserving such objects implies not only conserving their material shell, but also reconstructing the context of their use, tracing the individuals, artifacts, institutions, disciplines, concepts with which they were associated. The contextualization of such items therefore seems essential for *preserving* them *as* objects of knowledge.

Another argument concerns the epistemological affordances of the digital archive of science. As commonly noticed, the digital turn has provided unprecedented access to a myriad of collections and artifacts from institutional holdings and repositories. Yet, digitization is not enough to integrate this new multitude of scientific objects into historical research. Making them data readable and interpretable for the history of science requires not only digitization, but also the *semantization* of digital collections.

With these arguments in mind, in this part, I will be exploring the possibilities for *semantizing* the residues of science within the digital archive. Lacking a more subtle and elegant term, I use "semantization" to articulate the range of questions about giving meaning to a thing: How to make sense of the artifact within the digital archive? How can artifacts be connected to meanings? What forms of contextualization are possible within digital collections? What forms of connections can be made?

"Semantization" also points to the method I am using – semantic modelling. I construct a semantic model (or ontology) for contextualizing the residues of science within the archive. The ontology provides a vocabulary for describing the scientific artifacts and outlines the potential relationships that can link archival items (and their digital representations) with each other, as well as with individuals, disciplines, concepts, and institutions. This approach essentially involves *modelling the archive*, imposing a certain schema upon it, and in such a way so that it would be able to produce new knowledge.

As I will elaborate further in the subsequent chapters, my objective is not to devise a singular, universal, and all-encompassing model for the archive of scientific residues. Nor is my intention to offer practical recommendations or guidelines to archival and scientific bodies. Instead, this "critical making" approach (Ratto 2011) seeks to explore the potential, constraints, and frontiers of archive modelling. It probes into how a scientific residue could be contextualized and interpreted, both within and by means of the digital archive.

Upon constructing the model, I bring it to life with a case study that involves modelling the "archive" of one particular scientific residue. A single case study on its own can hardly validate the proposed *semantization* perspective. Yet it puts it in action, giving us a glimpse of its epistemic possibilities. It serves to probe modelling as applied to the archive; to explore what can and cannot be captured by the model; to see what connections, interpretations and histories of a residue an "interpretive archive" can afford.

The residue in question is one particular device, or rather a family of devices, which was in use in psychological experiments of the nineteenth and twentieth centuries: namely, the *reaction key*.

3.1.1. The reaction key and its multiple contexts

In 1879[137], Wilhelm Wundt, credited as “the father of experimental psychology”, established the psychological laboratory at the University of Leipzig. In this laboratory, mental processes – such as perception, sensation, attention or will – were subjected to experiment and measurement. As a method of the “new psychology”, Wundt developed an experimental program that became known as “mental chronometry” or reaction-time research[138]. Under this program, Wundt and his students conducted a series of experiments “to measure the precise pace of the brain as magic lantern” (Canales 2009, 10). Measuring the time of the subject’s response to visual, auditory, tactile, and other stimuli, they sought to establish the duration of psychological processes (e.g., sensory perception).

Reaction-time research was inherited from physiological research, in particular that of Hermann von Helmholtz, the first to establish that nerve transmission takes time[139]. In 1868, another physiologist, F. C. Donders, introduced the so-called ‘subtractive method’ for measuring the timing of mental processes. He suggested that the reaction time could be divided into several stages, and that by offering different types of tasks, it was possible to divide these stages and count the time of each of them. Despite a number of subtle disagreements with Donders’ assumptions (Robinson 2001, 164-166), Wundt generally adopted his method and even believed that it would allow calculating the exact time of an entry of impression into consciousness.

Astronomy was yet another important source. The question of reaction time arose from an astronomical problem known as the “personal equation” – discrepancies in the observations of star transits by different observers. While trying to solve the personal equation problem, in 1861 Adolph Hirsch, head of the Neuchâtel Observatory, conducted his “Chronoscopic experiments on the the speed of various sensory impressions”. The experimental set-up for Wundt’s “Reaction experiment with Hipp Chronoscope” was adopted directly from Hirsch (Schmidgen 2003), but became the signature of the Leipzig laboratory and of the entire new discipline.

---

137 According to some accounts, the actual date of the establishment of the laboratory is 1875 (Harper 1950). On the establishment of the Leipzig Laboratory see also Bringmann and Ungerer 1980. For an early account of the laboratory’s research, see Cattell 1888.

138 On reaction-time research in Wundt’s laboratory, see Boring 1961; Coon 1993; Benshop and Draaisma 2000; Robinson 2001, Schmidgen 2003. On the history of reaction-time research in a broader perspective, see Canales 2009; Schmindgen 2014.

139 Wundt worked as Helmholtz’s assistant at the University of Heidelberg. On Helmholtz’s time experiments, see Schmidgen 2002.

In fact, as Wundt insisted, the reaction time experiments investigated purely psychological phenomena: the way ideas (or representations) enter consciousness. During the 1860s, Wundt began to elaborate a program of psychology as an independent experimental science, concerned with a distinct sphere of the psyche. In the 1880s, it was already in full swing at the Leipzig laboratory. In 1886 James McKeen Cattell, one of Wundt's first doctoral students, observed that "the relation of the sensation to the stimulus and the time taken up by mental processes are the two subjects in which the best results have been reached by experimental psychology. These results are important enough to prove those to be wrong who with Kant hold that psychology can never become an exact science" (Cattell 1886d, 63).

The institutional and intellectual landscape of and around experimental psychology in the late nineteenth century and early twentieth century was exceptionally dense. It was the time of the birth and institutionalization of psychology. By the 1890s, the Leipzig Institute was by no means the only one in Germany: there were also the laboratories of G. E. Müller in Göttingen, Carl Stumpf in Berlin, Hermann Ebbinghaus in Breslau, Oswald Külpe in Würzburg[140]. In parallel, William James' laboratory at Harvard was active in America from the 1870s. The second half of the 1880s onwards saw the exponential growth of psychological laboratories in the US, many of them opened by former students of Wundt[141], practicing his method, the experimental program and the (new) ethos of the scientific psychologist.

This whole milieu was charged with discussions, debates and controversies; Wundt's positions on various questions have been refuted time and again[142]. Both Freudism and Gestalt psychology were partly reactions to the experimental program for studying consciousness. Philosophers also engaged in a dialogue (or, more precisely, a permanent polemic) with the paradigm of experimental psychology; for instance, Wilhelm Dilthey, who articulated his concept of understanding, and Edmund Husserl, who framed his program of phenomenology against "psychologism". By contrast, the main representatives of American pragmatism, William James and Charles Sanders Peirce, were themselves actively involved in the formation of experimental psychology. Wundt had a great influence on the formation of anthropology as well, as both Franz Boas and Bronisław Malinowski studied under him for some time.

140 For an overview and statistics of all psychological laboratories prior to 1900 in the United States, Germany, France, and Britain, see Harper 1950, 160.

141 Here are only some of Wundt's students who have headed the American laboratories: G.S. Hall (Johns Hopkins), J.M. Cattell (Pennsylvania, Columbia), Münsterberg (Harvard), Titchener (Cornwell), Scripture (Yale).

142 For some examples of the debates, see Rieber 1980, Mülberger 2012, Hui 2012, 123-148, Russo Krauss 2017, 2022.

I will be looking at this piece of both experimental and intellectual history from a rather unusual perspective – that of a small thing, a minor device or accessory, a material trace, left behind in archival collections. The reaction key was an electrical switch used to capture the subject's response in the course of reaction-time experiment. The very first reaction key was an ordinary telegraph key, which in the psychological set-up changed its purpose: instead of transmitting a message, it recorded a reaction. Responding to a stimulus, the subject activated the key (i.e., pressed a button, released a telegraph key, opened his mouth to speak), which in turn broke the interconnected electrical circuit of apparatuses and stoped a clock (chronoscope, chronograph, timer) that recorded the time between giving a stimulus and receiving a response. The reaction keys hence quite literally connected the test subject to a complex laboratory set-up, being the only mediator between the subject's body (and mind!) and the experimental machinery. The key captured the subject's reaction, made it available for measurement, and thereby provided a peculiar form of access to the epistemic objects of psychological research, be they consciousness, will, apperception, or cognitive processes.

The history of science as well as media history have actively discussed early experimental set-ups in psychology, yet their emphasis has been either on time-measuring devices (chronoscopes, chronographs, timers and their role in measuring the subject)[143] or on devices for producing stimuli (tachistoscopes, memory drums, metronomes…)[144]. The history of the "minor" devices for capturing a response, however, has not yet been written.

Instead of writing this history, I craft it weaving numerous archival bits and pieces: objects exhibited in digital collections, leftover photographs, catalogs, and publications. The pattern of this weave, or to put it more prosaically, the logic of juxtaposing digital archival fragments, is to be defined within the framework of an experimental ontology.

3.1.2. The key and the house: semantic/hermeneutic vs antiquarian attitude

To frame my perspective, let me take a moment away from the reaction key and have a look at another residue: Wundt's summer home in Großbothen, not far from Leipzig. Currently it is being renovated, but prior to the reconstruction it was preserved as a digital panorama offering a 360-degree virtual tour[145]. In this tour (Fig. 3.1.1), the user moves virtually through the rooms

[143] For some examples of research, see Gundlach 1996, Schmidgen 2003 and 2005, Benshop, Draaisma 2000.

[144] On the history of the tachistoscope, see Benshop 1998, Crary 2001, Acland, 2012. On the history of the memory drum see Haupt 2001, the history of the metronome can be traced in Bonus 2017.

[145] [https://www.berlin-web.de/wundt/]

of the deserted house and abandoned garden, climbs the decaying stairs, looks at the crumbling walls and through the shattered windows. The house is ruined. Although one can see some traces of the life that used to be there – pieces of wallpaper, remnants of the fireplace, planks, friezes – there is no certainty that these are the traces of Wundt and not of the subsequent owners. So what does this digital representation of the ruin preserve and communicate?

Figure 3.1.1. Wundt's summer home in Großbothen. [https://www.berlin-web.de/wundt/]

The answer would be: the age-value (*Alterswert*), the emphatic experiencing of the traces of time. Alois Riegl, who coined the term, described the experience as follows:

> These monuments are nothing more than indispensable catalysts which trigger in the beholder a sense of the life cycle, of the emergence of the particular from the general and its gradual but inevitable dissolution back into the general. This immediate emotional effect depends on neither scholarly knowledge nor historical education for its satisfaction, since it is evoked by mere sensory perception. ... We will henceforth call this the age-value (Riegl 1982 [1903], 24).

Age-value manifests itself in the traces of time, "in the wear and tear of buildings and objects," (ibid., 32), in their patina praised by John Ruskin as "the golden stain of time" and their mold the antiquarian breathes, according to Nietzsche. Traces of time are seen as a bearer of the object's aura and proof of the object's authenticity in need of conservation.

This 'antiquarian attitude' (Bann 1990, 102), the imperative to preserve the thing as it is, formed the basis of an entire science of restoration. The (European) vernacular approach to historical preservation, set by the *Venice Charter* (1964), puts the notion of authenticity at the center of the heritage politics[146] and encourages preserving original and authentic materials and

[146] On the history of the debate on authenticity, see Starn 2002.

the traces of time imprinted upon them. This antiquarian attitude, the attention to the authenticity, the aura, the immersiveness of the experience of the past, is also perfectly applicable to digital archives, exhibitions, and installations. Archive (even the digital one) is associated with preservation, and far less so with modelling.

The approach I practice in this part stands in contrast to this antiquarian attitude. The emotional experience of the residue contrasts with its *interpretation*. The preservation of matter, seen as the guarantee of authenticity, is counteracted by the *modelling* of meanings. Making archival artifacts readable and interpretable means introducing them into a (historical, cultural, pragmatic) context, relating them to other artifacts and agents[147]. In this logic, an alternative to the virtual tour offering an experience of immersion in a ruin would be a *montage* or *collage* bringing together different testimonies about the house and its inhabitants. Instead of storing a historical fragment as it is, I thus advocate *preservation through remodelling*, establishing connections, combining and re-combining various fragments, remnants, and scraps of the past scattered throughout various collections and libraries. This sort of preservation requires certain craft, constructing, bricolage: clipping, carving out details of the past and piecing them together anew into some new collage.

Getting back to the key: consider, for instance, this representation of the lip key[148] exhibited by the Center for the History of Psychology at the University of Akron (Fig. 3.1.2). To perceive it through antiquarian attitude would mean to zoom in on the photograph, experiencing its cracks and fractures, the aging of the metal, and mold on the tips. The attitude is still quite applicable, although it works somewhat less well than for the house. In fact, any object showing traces of time could be approached in this way. Yet thematizing it as an object of knowledge or *making it readable* involves posing a great deal of questions that have yet to be answered: What is it for? What can it testify to? Who used it? What part did it take in an experiment? What did the subject feel while holding it between his lips? What was its mechanism? How did it end up in the archive, was it important for some reason?

[147] In this sense, the approach described is in line with the material turn in history, social and cultural studies (Appadurai 1988; Lubar, Kingery 1995; Baird 2004)

[148] [https://collections.uakron.edu/digital/collection/p15960coll7/id/735/rec/2].

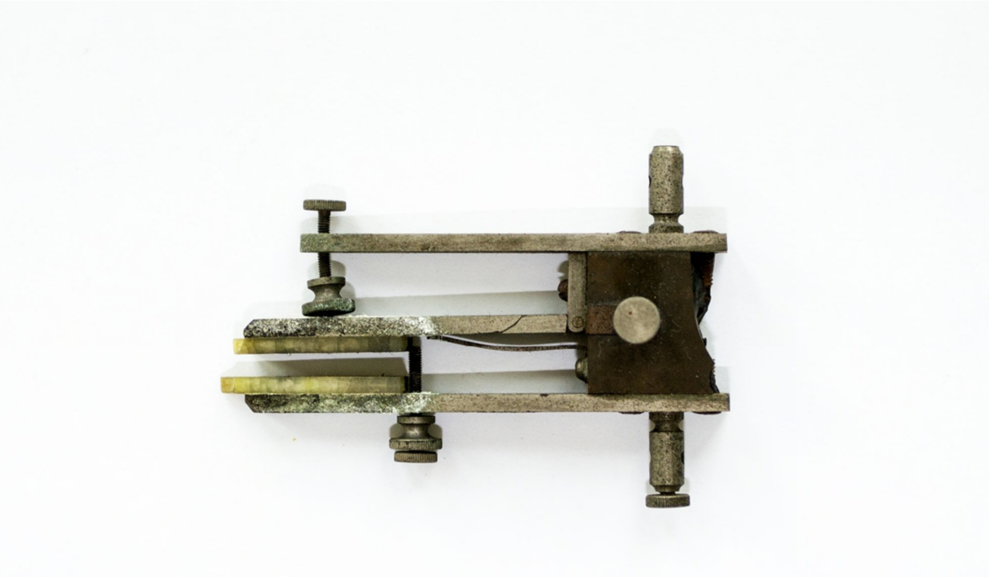


Figure 3.1.2. The lip key (Center for the History of Psychology at the University of Akron)

In this part, I advocate this particular (methodological/epistemological) stance toward the readability of the thing in the archive. This position in a certain sense carries on Don Ihde's project on *expanding hermeneutics*: "a unified and specialized mode of *thing interpretation*" (Ihde 1999, 8). As he puts it, "the postmodern hermeneutics of things must find ways to give voices to things, to let them speak for themselves" (ibid., 158). If things can testify, it would be worth learning to understand and interpret their testimonies. Ihde suggests the interpretation would arise "through imaging instruments" (ibid., 8), in my case the interpretation would be through digital archival representations.

Before turning to the interpretation itself, I will outline the method – namely, semantic modelling and ontology.

3.2. Ontology as a knowledge machine

> None of the programs that presuppose history as a sequence of situations can make the assertion '*History is a sequence of situations*'
>
> John McCarthy

In the context of knowledge representation[149], the discussion of ontologies began in the field of artificial intelligence in the 1980s. The term is believed to have been introduced to the field by John McCarthy (the originator of "artificial intelligence" as well), as a result of his reading of Quine (Smith, Welty 2001). In 1983, John F. Sowa defines ontology as "a catalogue of everything that makes up a world, how it is put together, and how it works... a catalogue of concept and relation types" (Sowa 1983, 294). Ontology thus fits into the range of "tools of knowledge" touched upon in the brief sketch on knowledge repository (chapter 1.1.2): cataloguing systems, indexes, paper slips, file cabinets and different sorts of "paper machines" (Krajewski 2011) that make data processing possible.

Already in the 1980s, reflection on ontologies developed into a breathtaking mix of philosophy, cognitive psychology, formal logic, linguistics, and computer science. For instance, Sowa begins his 1983 book with the philosophical basis (ch. 1) and psychological evidence (ch. 2), to then move on to reasoning and computation (ch. 4), knowledge engineering (ch. 6) and its mathematical background (appendix A).

In 1993 Tom Gruber gave the classic definition of ontology as "a specification of conceptualization" (Gruber 1993, 199). Both terms of this terse formulation are crucial: the 'specification' refers to a technical description (Gruber himself gives the example of a specification of a program), a formal language readable by a computer. 'Conceptualization' means that ontology describes some system of concepts and notions, *another representation*, rather than 'reality'. Ontology, in other words, is a formal representation of some concepts and the connections between them, within a domain (of knowledge).

Within the field of Artificial Intelligence, ontology is associated with certain epistemological assumptions and commitments. First, borrowing the philosophical concept of ontology to designate some form of *representation*, AI makes it explicit that for its systems, "what 'exists' is that which can be represented" (Gruber 1995, 907). Second, within the AI

149 Ontology and semantic models are used synonymously in this part, semantic modelling is understood as the process of creating and populating a semantic model. On ontology and its crossmeanings in philosophy, computer science, and information science, see Almeida 2013. For concept meanings within AI and knowledge representation, see Vickery 1997.

framework, any knowledge representation is understood as a surrogate, an imperfect copy, which is inevitably inaccurate and therefore can lead to incorrect inferences (Davis, Randall, Shrobe, Szolovits 1993). Third, knowledge representation implies ontological commitments – "a strong pair of glasses that determine what we can see, bringing some part of the world into sharp focus at the expense of blurring other parts" (ibid., 19).

In 2001, in an article in the *Scientific American*, Tim Berners-Lee remarked that "Knowledge representation… is currently in a state comparable to that of hypertext before the advent of the Web: it is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world (Berners-Lee, Hendler, Lassila 2001). So as a remedy, semantic technologies have been included in the Semantic Web stack of standards. The main task of ontologies became "adding logic to the web," while reflections on their philosophical and cognitive foundations took a back seat. The overall intent of the Semantic Web project (renamed as Linked Data in 2009) is to aggregate knowledge on the Internet by making it machine readable. As explained in the same article, "the Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (ibidem). To create a web of data, the World Wide Web Consortium (W3C), led by Berners Lee, developed a number of Web standards that are stacked into what is known as a Semantic Web Layer Cake (Fig. 3.2.1).

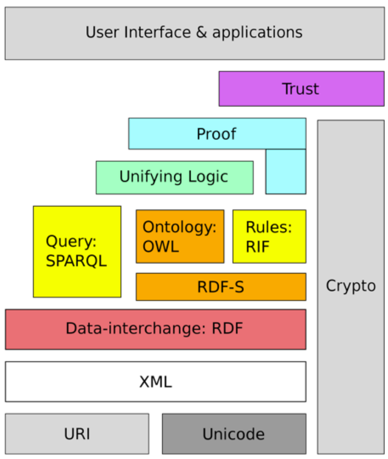


Figure 3.2.1. Semantic Web Layer Cake

The "cake" represents the established technological standards as, to quote semantician François Rastier, "the steps of a *gradus ad Parnassum*, which leads from *Unicode* to *Trust*" (Rastier 2010, 95). The Semantic Web relies on successive layers of standards, each one drawing on the previous layers. I will only comment upon some of the intermediate steps which are of relevance to us, leaving aside both "Unicode" and "Trust".

The basic unit of semantic technologies are RDF[150] statements representing information in the form of a logical primitives or a *triples* (subject—predicate—object). Any proposition within the Semantic Web is represented in the form of such simple expressions, in which the subject and the object are connected by means of a certain predicate. For instance, the proposition "Wilhelm Wundt founded the Laboratory for experimental psychology in Leipzig in 1879" can be decomposed into a few triples[151]:

> Wilhelm Wundt (subject) founded (predicate) Leipzig laboratory for experimental psychology (object)
> Leipzig laboratory for the experimental psychology (subject) was founded in (predicate) 1879 (object)
> Leipzig laboratory for the experimental psychology (subject) was situated in (predicate) Leipzig (object)

Every statement may be also represented as a graph (Fig. 3.2.2.).

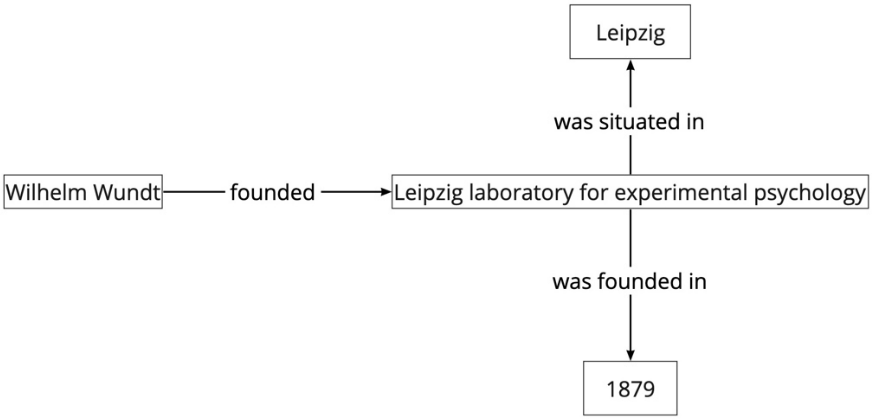


Figure 3.2.2. Graph representation of the triples

Further, each unique subject, object and predicate has its unique identifier (URI), which allows to identify them on the Web. That is to say, the *Wilhelm Wundt* of the example above will correspond to the *Wilhelm Wundt* in Wikipedia, DBpedia, and numerous other databases: all the information about him (triples in which the entity *Wilhelm Wundt* takes part) will be accumulated in the web.

The categories (or *classes*) of objects and rules, according to which they can be linked, are set by ontologies. This brings us to the OWL level: a language standard for creating ontologies, produced by the consortium. Ontology defines the nomenclature of classes to which instances belong, and the rules of relations between these classes, according to which, the

---

[150] RDF stands for Resource Description Framework.

[151] The proposition can be decomposed into triples in different ways. The proposed version is only one of the possible options.

model can make logical inferences. Returning to the example with Wundt: if the rule in ontology specifies that only a person can *found* something (or, more precisely, that the predicate “founded” can only be assigned to the instances of the class “person”), then the system itself will draw the logical conclusion that *Wilhelm Wundt* is a person. This *reasoning* power of ontology – the ability to draw logical conclusions, according to given rules – implies that the semantic model not only reflects the information that is encoded into it, but is also capable of *producing* some new information.

But besides that, ontology provides, for sure, a wide range of possibilities for search, which is done in Semantic Web by another language, called SPARQL. Ontology makes it possible to perform searches for entities, relations, classes, separately and in various combinations: one can, for example, find all that Wilhelm Wundt founded, or all the facts of “foundation” (all the triples with the predicate “founded”), or all the institutions located in Leipzig, or all that happened in 1879.

3.2.1. Ontology in science and its history

The landscape of existing ontologies is so vast that any attempt to outline it would be doomed to failure. I make only a few very general remarks, which are relevant to the model I propose. To begin with, there are a number of ontologies that have become the unequivocal standard for modelling a particular field: e.g. the GeoNames for locations, the FOAF for describing (social) relationships, the Dublin Core, BIBO for bibliography, SKOS for describing classifications and vocabularies, CIDOC CRM and several of its extensions serve as the standard for museum and archival descriptions, Schema.org for web pages, etc, etc.

Becoming a standard here means that the patterns of classes and rules established in these ontologies, as a rule, are borrowed when constructing new ontologies. Adopting and aligning ontologies, making compatible ontologies, is considered to be good practice. Numerous recommendations for creating ontologies advise using existing ontologies rather than inventing new ones. In other words, the representation of knowledge builds on already existing patterns of knowledge representation. Here again we are faced with the metaphor of a layered cake (of ontologies): each model is made up of (prior) models and standards and in turn is imported by new models.

Scientific ontologies make up a separate, immense and self-contained field, particularly in the life sciences[152]. Such domain ontologies serve primarily as tools for structuring and

[152] For overviews of the field of life science ontologies, see Schulze-Kremer, Barry 2005; Hartung, Kirsten, Rahm 2008; Panzarella, Veltri, Alcaro 2023.

describing scientific data, such as DNA sequences or cell types. In that sense, they can be seen as sequels to the data collection, systematization, and storage practices of science, such as the botanical plant classifications, or astronomical compendiums[153].

Numerous cross-disciplinary ontologies are also available to describe different types of scientific practice: scientific publications (CISP) and scientific discourse (ScholOnto[154], SWAN[155]), scientific experiments (EXPO[156]). The main purpose of such ontologies is to formalize scientific knowledge: for example, EXPO is presented as a model for formalizing experiments. The argument goes that the unproblematic, unambiguous language of ontology is more suitable for the representation of knowledge, than natural language. After all, ask the creators of EXPO, "is writing a scientific paper closer to writing poetry or a computer program?" (Soldatova, King, 801). Moreover, the authors of EXPO further argue that a unified formal description of experiments makes it possible to identify errors in conducting and outlining experiments. As an example, the authors annotate, via EXPO, two arbitrary scientific articles and conclude about their consistency and validity on the basis of the model. Thus, ontology no longer appears as merely a form of representation, but as a standard of judgment, a transparent and universal technology for the production of scientific facts.

Digital Humanities has been guided by the example of the life sciences since the late 2000s[157], considering the applicability of semantic modelling to the humanities, and in particular digital history. The examples of historical and history-of-science models are quite numerous[158], yet as a rule, they are destined for specific projects. None of them seems to have become the standard: either because historical reality is less amenable to formalization than scientific experiments, or because the humanities are more sensitive to the perils of creating a "universal language".

3.2.2. Critique of ontology

Despite the great variety of ontologies and initiatives, the Semantic Web utopia has not yet materialized. The Semantic Web project has been criticized in terms of technologies, business and publishing models, user policies, security issues, feasibility, sustainability and political

153 On archival practices of sciences cf. Daston 2017.
154 Buckingham, Motta 2000.
155 [https://www.w3.org/TR/hcls-swan/]
156 Soldatova, King 2006.
157 See, for example, Zöllner-Weber 2010; Jansen 2019.
158 See, for example, Meroño-Peñuela et al. 2014, Ide, Woolner 2007, Allen 2013.

decisions[159]. Yet its most profound problem from my point of view can be formulated as follows: by setting ontologies as a technological standard designed to ensure universal compatibility at all costs, the Semantic Web has adopted a somewhat simplistic conception of knowledge. This point was articulated in particular by François Rastier who accused the Semantic Web of "naive realism". In his view, the Semantic Web non-reflexively reproduces old principles of logical positivism: the idea that "extraction or representation [of knowledge] never modifies its content" (Rastier 2010, 96) and that chosen categories of description are context- and language independent. This is exactly the line of reasoning we have found in the creators of the EXPO ontology, who frame semantic models as a universal language, transparent, independent of context, and capable of finding contradictions in "reality".

The epistemological commitments made in the 1980s, and in particular the crucial thesis of ontology as representation, have been virtually neglected. The Semantic Web has shifted the focus from knowledge representation to information aggregation. Theoretical models have been replaced by technical standards designed to ensure the compatibility and interoperability of various description schemes. In the grand utopia of the semantic web, all the bits and pieces described in different ontologies are to come together in one gigantic network of knowledge, where everything is connected to everything, and any new piece of knowledge can be immediately woven into the existing web.

The pursuit of universal standards, however, obscures and puts into brackets one crucial dimension which is the transiency or historicity of any model. A model of a domain, as seen by its most expert representatives, reflects but some *current* consensus[160]. Even having exhaustively and universally described some field of knowledge, one cannot declare the task accomplished, for not only the data gets obsolete, but also the very *language of its description*.

The problem is reminiscent of Arthur Danto's thought experiment with the Ideal Chronicler, who in real time perfectly records everything that happens the way it happens (Danto 1968). Would his chronicle become the perfect historical narrative? Danto's answer is that the ideal chronicler would not be able to produce a single historical assertion. A historical assertion requires a certain distance between "reality" and itself. The chronicler, then, could neither compare one event with another, nor analyze it in the perspective of future events, nor recognize and establish its significance. As such an ideal chronicler, the Semantic Web seeks to capture and accumulate all the currently available representations, but fails to take into

[159] For a detailed review of the varied arguments of Linked Data critics, see Hogan 2020.

[160] One can imagine that for history (of science) this consensus, expressed and embodied in the language of ontology, would itself be an exceptionally interesting object of study.

account the fact that knowledge is not only aggregated, but is also changing, being revised, experiencing "revolutions," "paradigm shifts," and so on.

Out of these observations stem some of my own (epistemological) assumptions and commitments, which then formed the basis of my experimentations with semantic modelling:

First, I am not using semantic models as an end in themselves, nor do I seek to propose a universal model for the ontology of the archive of scientific residues. Rather, I approach semantic modelling as a technique for generating connections (or a specific version of the knowledge machine), which potential or usability for the digital archive I seek to elucidate. To put it differently, designing an ontology for me is not a solution to the questions posed, but rather a mode of investigation: *critical making* (Ratto 2011) or *research through design* (de Mourat et al. 2015).

Second, an ontology is a model, a representation, a sort of analytical apparatus, and this very fact needs to be explicated. An ontology neither corresponds to the world nor reflects it, but only offers a specific model of it, which has certain heuristic powers.

Third, an ontology, and the language(s) it offers for describing objects or events, are transient and reflect only the *ongoing* understanding of a particular domain. This fact must also be recognized, explicated and acknowledged within the very ontology.

These rather abstract propositions are embodied in the architecture of the ontology for scientific residues I propose. The most general structural idea is that instead of one universal model for describing data, an ontology should offer a number of such models, with each model being described, historicized, and referenced to the relevant literature. This plurality does not let any of the models to be taken as the only objective way of representation. In this way the data can to be treated as something not objectively given, but constructed, as *capta* (Drucker, 2011).

Within the proposed ontology, I established at once *three models* for describing the scientific residues: (1) "biographies", (2) "assemblages", (3) "mediations". Each of these models defines a specific vocabulary of description (i.e. classes of objects and rules of their relations) associated with a particular interpretative approach to (scientific) things. So, the first model follows the biographical approach to things articulated in late 1980s anthropology; the second one is drawn from assemblage and actor-network theory; the third one is informed by some notions from historical epistemology and techno-hermeneutics. Each of the models thus poses different questions and contextualizes the thing somewhat differently within the archive.

Further, each model is described in the ontology through its affiliation with certain researchers and an approach, a theory, a school of thought, and is grounded in certain publications (Fig. 3.2.3).

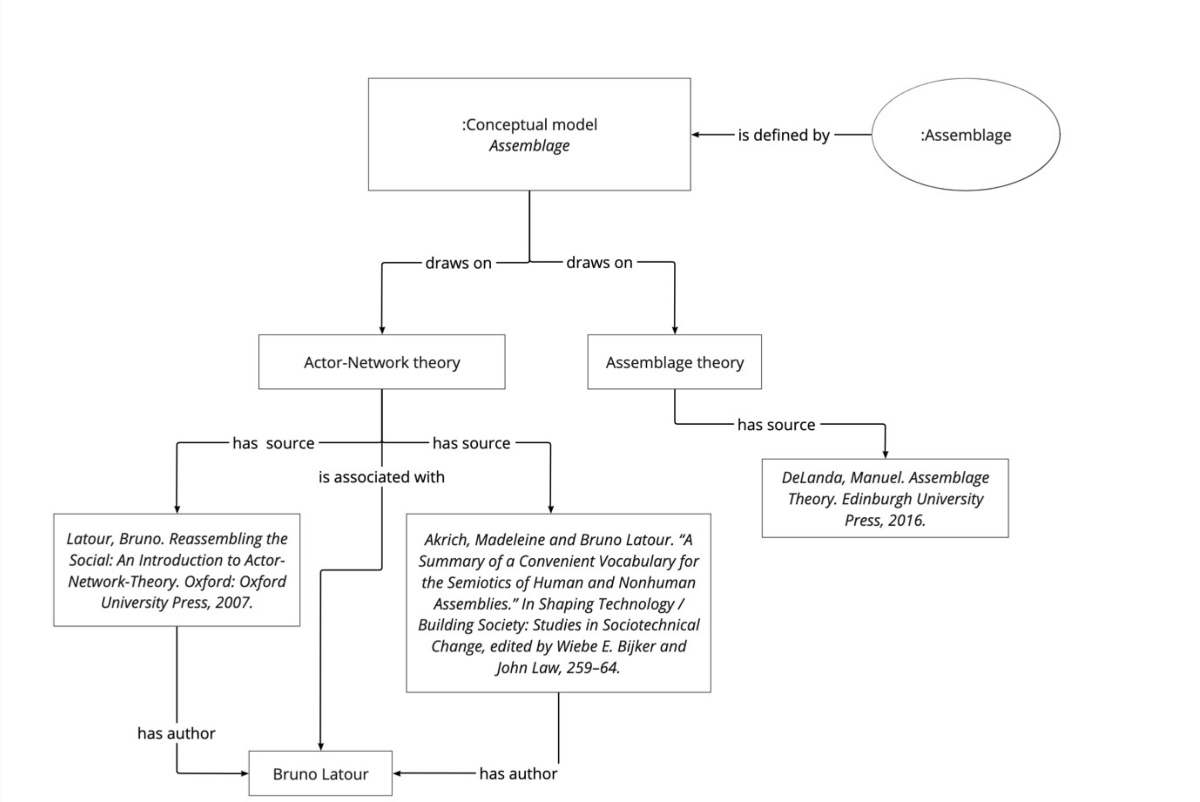


Figure 3.2.3. A rough sketch of the description of the "Assemblages" model in the ontology. The model draws on certain theories (approaches, vocabularies), which in turn have sources (publications). It can also be associated with people (as in this example), but also institutions, concepts, artifacts. The model, in turn, defines (describes) certain classes and properties (in this case, the Assemblage class).

Thus, the model of description gains historicity. Rather than an abstract and universal system of concepts, it turns into a schema associated with a certain research perspective, formulated in a certain context, which can be traced through a number of publications, which has its opponents and critics, as well as its own boundaries.

Each of these models, accordingly, implies different ways of making sense of scientific artifacts, setting a certain logic for making connections between them (such as biographical connections or relationships within an assemblage). Through the lenses of these models, I plotted the histories of the reaction key – all its connections, contexts, interpretations that could be reflected in the digital archive. But before I get to these histories, I need to detail the way the modelling was conducted.

3.3. Research through modelling: some preliminary notes

Willard McCarty suggests seeing modelling as "a form of craftsmanship set into the context of scholarship" (McCarty 2005, 22). If this is indeed the case, then some preliminary discussion of the technology of this "craftsmanship", is necessary before proceeding to conclusions, interpretations, and arguments. Hence the need for this purely 'procedural' chapter, recalling the descriptions of the experimental set-up in scientific articles. In it, I will outline how the modelling process was arranged, so as to move on to the models of description and the history of the reaction key in the coming chapters.

The process was broken down into a number of sequential (albeit interrelated) steps: (1) designing of the ontology, (2) populating the ontology (with some fragments of the reaction key history), and (3) making an exhibition to illustrate some of the conclusions and tendencies made possible by the ontology.

This brief chronology already makes it clear how different my "algorithm" of action is from the usual course of historical research, which generally begins in the archives and proceeds by approaching archival sources with a suitable/fruitful language of description. Instead, I impose a ready-made language or structure of concepts in advance, in which the "raw" historical material is further arranged. (In practice, these two processes were not so rigidly separated: some points of ontology remained empty, no matter how hard I tried to fill them; and some sections of ontology arose only when I lacked the means to describe certain fact). Now, let me give some more detail about each of the steps.

3.3.1. Designing the ontology (a workflow)

The process of ontology design, in turn, can be broken down into a few sequential stages (in practice, they were still a movement back and forth).

1 Delineating models of description

First, having undertaken quite a traditional literature review, I summarized some of the interpretive approaches to (scientific) objects and delineated the three models of description: "biographies", "assemblages", and "mediations", respectively. This list is by no means intended to be exhaustive: one can easily think of many other models for interpreting things (and they can even be easily attached to the current ontology). My choice was based rather on the limitations of the archival representation, as articulated in the preceding chapter. Each of the approaches I opted for offered a different form of contextualization of the scientific residue.

In each of the models I made no attempt to conform strictly to one theory, but rather combined and merged various theories and concepts quite freely: so, for example, I formulated the second model on the basis of various conceptions of assemblage, while in the "Mediations" model I brought together the concepts of Hans-Jörg Rheinberger and Don Ihde. Operationalizing different approaches to interpreting things, I did not aim to refine theories through ontology. Instead, through models and theories, I sought to ask questions to the residues of science.

2 Operationalization

The next stage was to translate each of the "models" into the language of ontology, that of entities, classes and the relations between them. I will discuss the particular concepts and their modelling in the subsequent chapters, and for now limit myself to a few general remarks.

The process of translation in practice involves (1) creating a nomenclature of categories, (2) defining (both possible and not possible) relations between them, and (3) establishing a system of description (annotation) of both categories and their relations.

(1) Establishing nomenclatures of categories implies figuring out what kinds of entities - people, things, concepts - are to be described in the model and how they relate to each other. For example, how do *people* and *artifacts* relate? Does one constitute a subtype of the other? Or do they have to be unified by some other class? In addition, the nomenclature involves defining a vocabulary or system of naming: e.g., how to name the class of people or artifacts? One can imagine many different possibilities, each one involving a particular stance: *agents*, *actors* and *actants*, *people* and *things*.

(2) Defining relations between classes means developing a system of predicates (or *object properties*). For each such property, one has to define which classes it can relate to each other. For instance, what connections can be made between a person and an institution (representatives of presumptive classes *person* and *institution*)? A person can work or study in an institution, however, the reverse is not valid. All such connections and logical restrictions need to be specified in ontology.

(3) A system of annotation determines how classes and their relationships are to be characterized in the ontology: e.g., how to specify different kinds of classes, how to align certain entities with existing databases and archives, which identifiers to use, which classes require references to the literature. In my ontology, whenever possible, concepts and relationships are defined based on the models of description: so, for example, the "Mediation"

model defines the object properties *mediates/ is mediated by* and the class *Epistemic object*, which are in turn conceptualized in the relevant literature.

On the whole, translating into the language of ontology is a rather non-linear process, filled with choices, problems, and uncertainties. Some of these issues and ambiguities are outlined further below.

Degree of adherence to theories and approaches. Grounding the ontology in certain approaches to things, I was faced with the question as to what extent to adhere to the concepts and language of this or that approach. How to decide which concepts to carry into ontology and which not? The most telling example is Actor-network theory, which I brought into play in the “Assemblage” model. Madeleine Akrich and Bruno Latour have developed an entire analytical vocabulary (Akrich, Latour 1994). How and to what extent should it be adopted in ontology? This is one of those choices that completely depends on the will of the ontology maker. In my case, the choice was determined by the goals of the model and the questions posed by it. (In the ANT example, I borrowed some concepts from Latour, particularly the concept of inscription, but did not follow its entire vocabulary).

The problem of untranslatability. Some categories can technically be moved into an ontology, but cannot be really *captured* by it. Take, for example, the notion of epistemic objects introduced by Hans-Jörg Rheinberger (1997), which I use in the “Mediation” model. In ontology it can be defined as what is being studied (by means of some “technical objects”). However, when encoded, it still loses an essential part of the conceptual intent: the constant change, unfolding, redefinition of epistemic objects is hardly operationalizable within ontology (this is further discussed in the corresponding chapter on “Mediations”).

Performativity of naming. The designer of ontology has the power (or curse) to *name* things; and naming in ontology is nothing compared to making a reference in natural language. Within ontology, quite literally, naming something ‘an invention’ *makes* it an invention, and naming something ‘a reaction key’ *makes* it a reaction key. Ontology generates entities: everything named in it turns out to be (ontologically!) real. That is why the creator of an ontology (just as the demiurge) has to be exaggeratedly cautious about creating entities and attributing facts to them (the ‘invention example’ will be discussed further in the “Biographies” chapter).

Anachronisms and craving for objectivity. Another problem, again related to naming, is anachronisms. The choice of categories for describing historical reality requires tracing the historicity of the very categories of description. In practice, however, it is impossible to avoid anachronistic descriptions, nor is it possible to create an “objective” artificial language that would eliminate all the contradictions of natural language. Ontology is itself a historical product after all, which is why it seems all we can do is try to keep track of anachronisms and be sensitive and attentive to them.

3 Implementation

To conclude the discussion of 'ontology making', I should give a number of technical details, as demanded by the genre. The ontology was built in the Protégé editor developed by Stanford University. It reuses a number of standard ontologies, including SKOS (for modelling some relations between concepts), Geonames (for modelling locations), BIBO and DublinCore for describing bibliographic data. I also borrowed some of the predicates of the CIDOC model as part of the biographical model. Whenever possible, I aligned entities with the existing databases: Geonames identifiers were used for encoding geographic locations, and Wikidata identifiers for encoding people, institutions, journals. I identified the archival objects by the identifiers from the collections in which they are located. For naming the digital archival representations I made use of their URLs.

The resulting model contains 130 (sub)classes and 273 object, data and annotation properties[161].

3.3.2. Populating the ontology

Sources and documentation

Turning to the history of the reaction key and its modelling in ontology, I will start with the selection of sources and some of their limitations.

For one thing, I have been following primarily the German and American traditions of experimental psychology, without, in fact, engaging materials from the French, British, Russian, Japanese and other national traditions. This choice was set by the very dynamics of the discipline development: by 1900 France and Britain had only two experimental psychology laboratories each, compared to twenty-five laboratories in America and ten in Germany (Harper 1950).

For another thing, and even more importantly, I have deliberately limited my sources to digital representations, i.e., those sources that are available online. Since this case study is meant to test the possibilities of semantic models for the digital archive, rather than to write a "true" full-scale history of the reaction key, I needed to limit the scope of the study. Since I am

[161] The ontology is available at [https://zenodo.org/records/10548179]

dealing with the digital archive, it seemed somewhat natural to engage mostly digital sources: i.e., those that are already available for "use" in the digital collections. As such, the history of the reaction key set forth in the following chapters is based entirely on digitized and publicly available materials.

Three major types of sources can be distinguished: digital collections (particularly, in the history of psychology)[162], digitized catalogs of instruments[163], and research publications in (experimental/ cognitive) psychology[164]. Each proposition reflected in the ontology has been documented by one of these three source types (for the most part, by publications). I also set out to review and collect, to the extent possible, all the representations of the reaction keys available in the digital collections. In addition, I traced other forms of the reaction key representations: photographs, images in catalogs and textbooks. The ontology thus contains a variety of types of evidence and puts them together in a certain way, in accordance with the rules outlined in the models.

Ontological disambiguation: type-object-representation

As noted repeatedly, semantic models describing a particular fragment of the world requires a logically consistent and hierarchical nomenclature of things. One unforeseen yet crucial problem I encountered while populating the ontology was that such a nomenclature of things does not correspond to the way we talk about things in natural language. As an example, consider this excerpt:

> *James McKeen Cattell invented the lip key in 1886. It was used in reaction-time experiments to record the subject's vocal response. The lip key can now be viewed on the website of the Harvard collection of scientific instruments. It consists of metal plates with ivory tips being placed between the subject's lips.*

Translating this text into triples, one would need to relate, through a variety of properties, some entity "lip key" to other entities: J.M. Cattell, materials, experiments, the Harvard collection.

---

[162] Collections on the history of psychology include: Wundt's archive at Leipzig University; Archives of the History of American Psychology at the University of Akron; Harvard collection of historical scientific instruments, The collection of psychological instruments at the University of Würzburg, Belgrade collection of scientific instruments, University of Toronto scientific instruments collection. In addition, I have used the Internet archive, which stores and displays reaction time software.

[163] A great number of psychological instrument catalogs from the late 19th and 20th centuries have been digitized by the "Virtual laboratory" project (Max Planck Institute for the History of Science). The "Virtual Laboratory" was created in the early 2000s and remains the premier and most invaluable online resource for the history of psychology and physiology. It is accessible via the link: [https://vlp.mpiwg-berlin.mpg.de/index_html]. For more on the project, see Schmidgen, Evans 2003.

[164] For the publications I used the Internet Archive as well as the American Psychological Association (APA) databases.

The problem, however, is that these properties actually describe different "lip keys". One and the same "lip key" in the text in fact refers to a number of entities: 1) a type of instrument (which, for example, was invented, but which cannot be seen in the Harvard collection!); 2) a concrete tangible thing, a realization of the type (which consists of some materials, was produced by some manufacturer and is located in the Harvard collection); 3) a digital representation of that particular thing (which cannot be said to consist of some materials, but which is represented on the Harvard's website).

Ontology must keep track of all these different entities if it is to capture all these specified properties. Thus, encoding even such basic 'facts' requires a peculiar procedure of ontological disambiguation, clarifying the types and sorts of things. Quite a lot of time and effort went into this procedure[165]. The solution I formulated is as follows (Fig. 3.3.1):

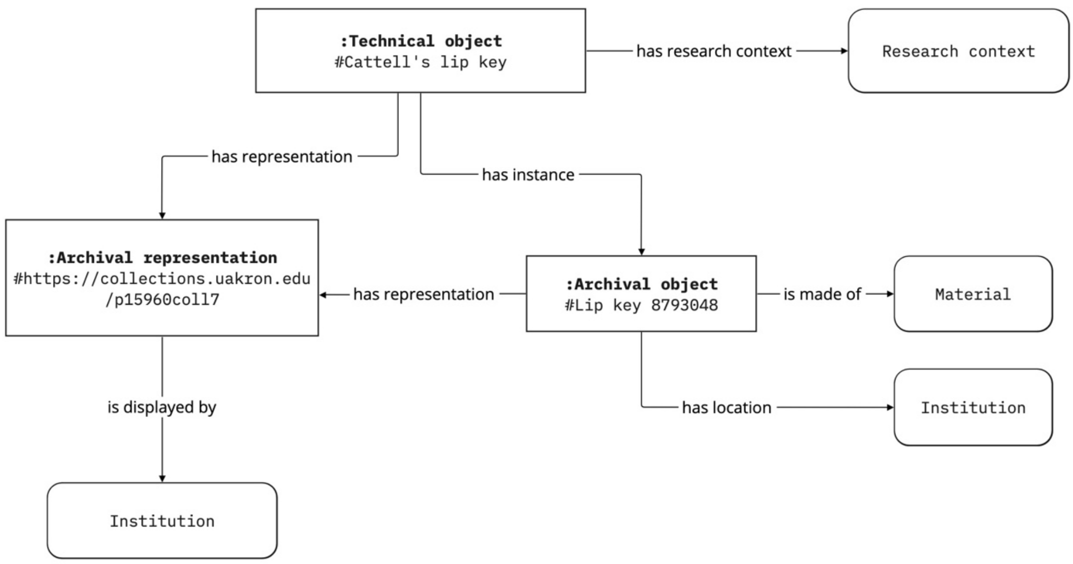


Figure 3.3.1. An example of lip key representation in the ontology: technical object (or type of device), archival object, archival representation. Each of these entities can be associated with its own set of properties.

In the ontology, entities are hence described through a triad: a *type* (e.g., of an instrument), an *archival object* (a tangible thing stored in a physical archive), and a *digital representation* of that object. In addition, the *archival object* must be related to the institution (collection) in which it resides; and the *digital representation* must be related to the archival object it represents.

---

[165] This example shows quite well yet another problem: even if the ontology is based on some pre-defined approaches, still its designer is forced to make a series of 'ontological' (as well as epistemological) decisions that go beyond these approaches.

All members of the triad are related to each other. In this way, each of the modifications of the reaction key is associated in the ontology with concrete implementations stored in archives and with their representations in digital collections.

Querying the ontology and making conclusions

The populated ontology can be used to make queries. All the bits and pieces arranged in its cells and shelves can now be retrieved in various combinations. In this way it specifies and produces a variety of connections between entities, which can then be reflected in the digital archive. Ontology thus sets the logic of the juxtaposition of archival artifacts, which I endeavored to represent in the digital exhibitions "Key2Mind".

Further, it makes it possible to perform distant reading and quantitative analysis of data, to identify patterns and trends. Although each of the models was designed to answer certain questions, at times unexpected patterns did come to light, calling for further interpretation.

In some cases, in contrast, certain interesting observations have arisen when particular points and clauses of the ontology have remained *unpopulated*. Sometimes it was indicative of some characteristics of the model itself (and made me rephrase the question and re-model the model), or other times it was suggestive of some kind of historical circumstances to look at more closely.

Moreover, some patterns became evident not even from the queries, but from the very process of populating the ontology, which appears as a very specific form of research. Thus, for example, by following archival representations as required by ontology, I came to discover that reaction keys pass from experimental psychology to cognitive psychology (which will be discussed in detail in the "Biographies" chapter).

3.3.3. Exhibiting reaction keys

In order to illustrate some of the conclusions and trends that the ontology brings to light, I have made a virtual exhibition called *Key2Mind*[166].

The exhibition reflects only part of the inferences and combinations offered by the ontology. It does, however, offer the advantage of exhibiting the meticulously collected fragments of the history of reaction keys: archival representations, images and text excerpts clipped from catalogs, inscriptions taken from publications, old photographs, reaction time software from the 1980s… The exhibition places all these leftovers in a certain order prescribed

[166] The exhibition is available via the link: [https://key2mind.omeka.net/]

by ontology. In it, the artifacts are contextualized through other objects rather than through categories and linear descriptions. In this sense, the exhibition succeeds in what the text of this very part fails to do: it shows the connectivity of archival artifacts by arranging, positioning them next to each other in different combinations.

That said, the exhibition also has very significant limitations, both as compared to the ontology and as compared to the text. First, it reflects only a small fraction of the combinations that the ontology makes possible. Second, the possibilities for representing the reaction key histories were constrained by the technical *dispositif* of the *Omeka*[167] platform.

Launched by the Roy Rosenzweig Center for History and New Media, *Omeka* is a platform for creating, storing and exhibiting digital collections. It is used extensively by institutional archives, in particular, those exhibiting the collections of scientific residues. *Omeka* offers a ready-to-use interface with the possibilities of structuring metadata and tags, but at the same time significantly limits the ways in which artifacts can be juxtaposed, connected, brought together. As such, it is quite telling (and confirms my findings from Part II) that the platform for digital collections provides opportunities for display (*visibility*), but not for making links and relating artifacts (*readability*). In practice, this meant that I still had to add text and timelines to the exhibitions. This limitation is probably particularly noticeable in the "Assemblage" model, in which reading across the assemblages, in fact, could not be exemplified.

Nevertheless, *Omeka* made it possible to literally *create* and assemble the archive around the reaction key. In that sense, this research turns out to be not only through modelling, but also through archive-making.

The remainder of this part goes on to present the three models and the histories of the reaction key. Each of the models – biographies, assemblages and mediations – is addressed in a separate chapter. The chapters are all arranged in the same way: each begins with a brief summary of a theoretical approach, then looks at the operationalization of that approach in the ontology, before moving on to some observations and episodes from the histories of the reaction key.

167 [https://omeka.org/]

3.4. Biographies

3.4.1. Thing biographies

The biographical approach to things takes shape in the anthropology of the 1980s[168], and eventually makes its way into archaeology (Gosden, Marshall 1999, Joy 2009), social sciences, the history of science, material studies and other disciplines. It derives from the genre of biography, representing a subject's life through the metaphor of a trajectory, a journey, a story, made up of a succession of events[169]. In literary studies, the biographical model has been well-described and hardly considered a method, yet when applied to (everyday) objects, it comes to be, as Adam Drazin puts it, a "critical methodology" (Drazin 2020, 62). The crucial methodological twist comes from the very idea that things (like people) can have biographies. This idea was coined in the articles by Arjun Appadurai (1986) and Igor Kopytoff (1986), who (albeit from different positions) postulated that the value of a thing was produced and experienced in the course of its production, circulation, exchange, and other *social* transactions. These various events and interactions – connections with people, objects, institutions – constitute the meanings of the thing, making its biography. Both Appadurai and Kopytoff addressed primarily the value of commodities, but their argument readily carried over to the entire repertoire of relationships to/with/through objects well-known to anthropologists: a gift, a totem, a talisman, a memorial site, a souvenir.

The biographical approach suggests that things have 'social lives': they participate in social relationships, accumulate meanings, and, along with other agents, form social environments. As Kopytoff noted, the questions to be asked in the cultural biography of a thing are more or less the same as those posed in the biography of a person:

> What, sociologically, are the biographical possibilities inherent in its "status" and in the period and culture, and how are these possibilities realized? Where does the thing come from and who made it? What has been its career so far, and what do people consider to be an ideal career for such things? What are the recognized "ages" or periods in the thing's "life," and what are the cultural markers for them? How does the thing's use change with its age, and what happens to it when it reaches the end of its usefulness? (Kopytoff 1986, 66-67).

Thus understood, the biography of things stands in contrast to the custodial (or provenance) history as practiced in the archives[170]. Provenance is intended to attest to the authenticity of

[168] For an detailed discussion of the approach see Hoskins 2006

[169] On the conventionality of these notions and the critique of biography as a method in the social sciences, see a classical essay by Pierre Bourdieu (1986).

[170] In the context of archaeology, Chris Gosden and Yvonne Marshall (1999) note a similar distinction between biography and use-life approaches.

things by offering an "objective" account of their production and possession. Biography, by contrast, invites one to trace not the credibility of a thing, but rather the ways in which it is given meaning in different contexts. While provenance tends toward formalized and "objective" description, biography does not: the same object or type of object can have a number of biographies, illuminating its trajectories in different contexts (including the archival one!). So, Samuel J. M. Alberti (2005) shows that an object's biography does not end with its entry into the archive and offers tracing the biographies objects before entering the museum, within the museum, and during its encounter with the viewer.

In the history of science, the cultural biography of things has received a good deal of attention and found quite fruitful ground. In addition to the mentioned work by Alberti, inspired directly by the anthropology of things, one finds also a number of studies tracing the life-trajectories of particular scientific artifacts (such as scientific instruments[171] or photographs[172]). In addition, the history of science has developed a distinct genre of scientific (or epistemic) objects biographies (Daston 2000) showing that even phenomena under study and theoretical entities such as electrons (Arabatzis 2005), the ether (Buchwald 2000) or cytoplasmic particles (Rheinberger 2000) have their (meaningful) life histories.

3.4.2. Operationalizing thing biographies

The biographical approach poses and answers rather simple questions: What are the "stages" in the life of a thing? What events was it involved in and how did it change through them? By whom and in what context was it created? In what (social/institutional/scientific) relationships has it taken part? How and at what point does it come out of use? The semantic model is thus aimed at tracing histories of production, circulation and usage of an object, its institutional and disciplinary settings, transfers between contexts, transformations, modifications, associations with individuals and institutions.

In the ontology, the biographical drive is fulfilled, above all, through the fact that artifacts (and other *agents*) engage in *events*. The event at this point is conceived both as a biographical (1) and ontological (2) category:

1) A biographical event is a major constituent of any life history. In literature, it is usually understood as that which changes the current state of affairs, separates one "stage" of life from

[171] For examples of such studies, see Pantalony 2011, Forstner, Walker 2020.

[172] E.g., Chadarevian 2003.

another[173]. These are birth and death, displacements in (cultural/social) space, transitions (from one condition to another), life-changing encounters. The list of these typically important events in a person's biography is set by prevailing social conventions (Bourdieu 1986): in the case of a person's biography, it is usually different variations of the pattern "birth – study – marriage – work – death".

In the ontology, I sought to provide a similar 'standard' outline for the life course of a scientific residue. The semantic model makes room for some typical events of the technological object's biography, such as *design*, *modification*, *production*, *relocation*, *exchange*, *transmission* (e.g., from one research context to another), *repair*, *archivation*, *discard*. The list of possible events could (and should) be expanded according to the case study.

2) Ontological event[174]. In the context of semantic modelling, the event is regarded as a special mode of ontology construction – event-based ontology. Arguably the most well-known example of such an approach is the CIDOC Conceptual Reference model adapted for describing and preserving cultural heritage, in which "the past is formulated as events involving 'persistent items'" (Doerr, Kritsotaki 2006, 58). Technically, this means that the basic unit of modelling is the class of events; it is through events situated in space and time that the so-called "persistent items" – persons, artifacts, ideas – come together. Interestingly, this logic is quite consistent with the biographical model, although the CIDOC CRM vocabulary itself aims at describing the provenance of an object rather than its movement through different contexts. In my ontology, besides changing the vocabulary (i.e., types of possible events and their participants), I have also slightly adjusted the specification of the *event* class. The event *must* be documented by some kind of evidence, be it publications, documents, images, oral testimonies. Therefore, even in cases where its exact date is uncertain, it can be defined by a special predicate *documentationDate*, which fixes the date of the event documentation (Fig. 3.4.1).

173 The literature on biography and the biographical event is remarkably vast. For some summary of the biographical theory in relation to the other disciplines see Renders, Haan 2014.

174 For an overview of the event-based ontologies cf. Ali et al. 2022.

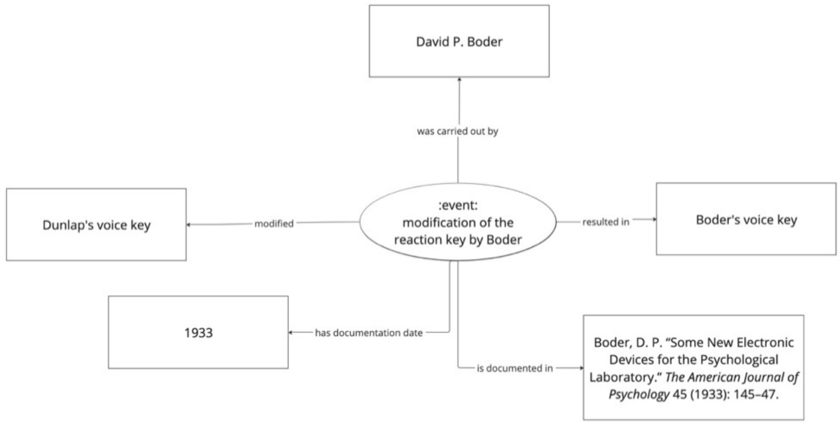


Figure 3.4.1. Encoding of an event in the ontology. The event of modification brings together different agents (the key that is subject to modification, the new version, and the designer) and is documented in the publication.

Apart from events, the biographical model also thematizes the typology of the agents engaged in the events (Fig. 3.4.2).

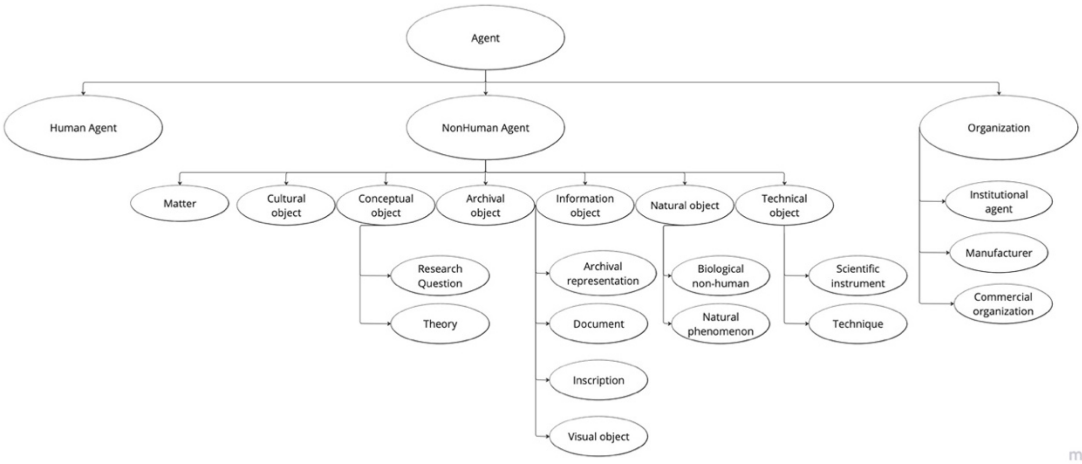


Figure 3.4.2. Structure of the Agent class.

The *agent* class is divided into three subgroups: *human agent*, *non-human agent*, and *organization*, which in turn are each divided into various subclasses. The most complex of the subclasses is that of *non-human agents*, which in turn is split into *technical object* (e.g. *scientific instrument* or *technology*), *information object* (*document*, *visual object*, *archival representation*), *archival object*, *natural object*, *cultural object*, *conceptual object* (e.g. *theory* or *concept*) subclasses[175]. This typology is by design open to change (all kinds of entities can be added to it) and does not claim to correspond to "reality": one and the same artifact can participate in multiple categories at once depending on a context.

Having described the categories of events and the agents, at this point we can move on to the history of the reaction key as it is framed through the prism of the biographical model.

[175] In subsequent models, *epistemic objects* and *inscriptions* are added as subclasses of the *agent* class, as discussed in the corresponding chapters.

3.4.3. Mapping diversity: inventions and modifications of the reaction keys in 1880-1900s

The only brief sketch of the history of the reaction key comes from Siegen K. Chou's article from the late 1920s. This history takes the form of an impressive list of the reaction key designs, constructions, modifications. With reference to (by then already classic) "Experimental psychology" by Edward B. Titchener, Chou records:

> Of voice-keys there are Cattell's, Roemer's, Wundt's, and Libby-Baldwin's; of eye-keys the eyelid-key listed by Zimmermann and the Dodge pendulum reaction-key; Cattell's lip-key, Meumann's biting-key, Kraepelin's speech-key, and Jastrow's speech-key are all adapted for vocal reactions; and Moldenhauer's stimulator is akin to a reaction key […]. The key ordinarily employed for this mode of reaction is the telegraph-key. Wundt had the early modified form; Zimmerman listed a simple form; Jastrow described the press-key; Ranschburg's key… (Chou 1929, 469).

This list provided a starting point for my modelling: following it, I began by mapping 'inventions' and modifications of the reaction keys. In modelling each particular design *event*, I referred to the publications and catalogs which documented it. Each such event related the model of the key to its manufacturer and creator (inventor or designer), and onward to his (academic) biography: the institutions he had worked and studied at, his research supervisors, and publications.

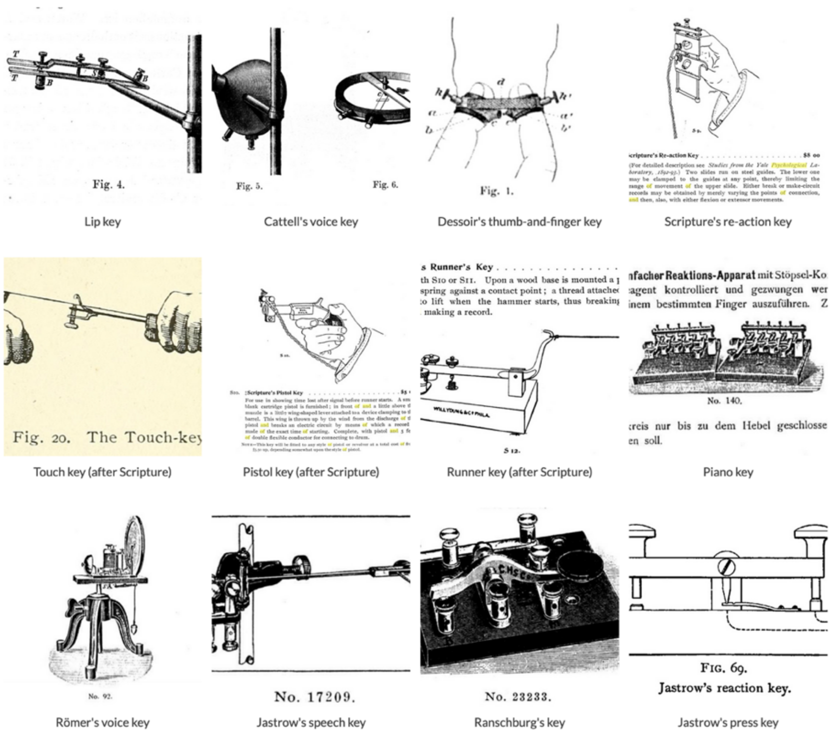


Figure 3.4.3. The reaction keys from before 1900.

In this way, a sort of biographical continuum was set in which the biographies of devices, their designers, manufacturers, and institutions intertwine. This continuum describes, primarily, the period from the 1880s to the 1900s, which appears as a time of constant designing, (re)inventing, and modifying of the reaction keys. During this time, the majority of the devices listed by Chou (no less than a dozen models) are being constructed (Fig. 3.4.3): the lip, speech and biting keys, eyelid key, various modifications of the voice key, a number of exotic keys by Edward Scripture, many variations of manual reaction keys, as well as the so-called key piano.

The key making began in Leipzig, with the first modification of the telegraph key by Wilhelm Wundt and his technician Carl Krille[176]. Throughout the 1880s and 1890s, Wundt's doctoral students constructed reaction keys for their experiments in the Leipzig laboratory and proceeded to do so as professors and founders of their own laboratories (particularly in the United States). So, for instance, James McKeen Cattell designed voice keys and lip keys back in Leipzig[177], before setting up psychology laboratories in Pennsylvania and Columbia; Edward Scripture constructed a great deal of reaction keys already in his laboratory at Yale University, and Joseph Jastrow developed his models at the University of Wisconsin-Madison. Further, most of the 'inventors' of the reaction keys had an associated manufacturer making the devices of his design. So Wundt collaborated with Carl Krille and later E. Zimmermann in Leipzig, Chicago's "Garden City Model Works" made Jastrow-designed devices, "Willyoung & Company" in Philadelphia manufactured Scripture's devices, and so on[178].

The social landscape of the *fin-de-siècle* key designers was indeed quite homogeneous: all but two[179] of them were prominent professors, founders of laboratories, and presidents of psychological societies. The reaction keys they designed bore their names: catalogs, textbooks[180], and academic publications (like the already cited article by Chou) identified them as *Cattell's* lip key, *Jastrow's* press key, *Scripture's* touch key, etc. The reaction keys of the period were authored, personalized, marked with the name of their creators.

Having outlined the general panorama of the "design events", I should say a few words about the specifics of modelling them. In the ontology, the process of creating the key is

[176] On Carl Krille's position as an invisible technician, see Benschop 1998.

[177] Cattell presented the devices in his dissertation (Cattell 1886a) as well as in an English-language article in *Mind* (Cattell 1886b)

[178] This tendency came to an end in the 1930s, when series production of instruments began.

[179] Two exceptions to the general rule were Ernst Römer, an assistant of Emil Kraepelin, and Julius Merkel, Wundt's doctoral student, whose trajectories left me with no evidence.

[180] One example is the already mentioned classic textbook by Titchener (1905), which has a whole section devoted to the reaction keys.

encoded as a *designing* event, with a subclass *modification*, which, in turn, is divided into several types.

The very categories I introduce here need some further discussion: in particular, the fact that I choose not to use the category of *invention*. It was a difficult conceptual choice to make. ‘Invention’ seemed to me a conventional form for describing the act of creating a new scientific instrument: in the history of psychology as well it is generally customary to speak of “inventions”[181]. Yet invention in the history of technology is by no means a neutral concept, it is flavored with numerous connotations (authorship, originality, uniqueness) and assumes a certain degree of novelty. In the semantic model, though, it is meant to serve as a class, a species, a model for certain facts, all of which must be equally adequate to it. That is, by ascribing a certain entity to the class of inventions, one does not simply describe it as an invention, but rather performatively sets it as a fact, models it as such. So, in the end I opted for a more neutral category of *designing*, which only captures a certain fact of the appearance of a new device.

For its part, identifying the distinctive “modification” type of event is primarily operational. It allows the ontology to capture those cases where a new model of the key results from reworking some particular prior model (as opposed to those cases where it is constructed without a known prototype). So, the *modification* event generates links between, for example, Dessoir’s thumb-and-finger key and Scripture’s key, showing that the latter is a modification of the former[182]. These cases of modification allow tracing some continuity of the devices: e.g, modelling successive revisions of the voice key: from Cattell’s early version to Römer’s modification to the cheap and popular model by Dunlap. Beyond that, however, the modification events enable one to keep track of the communication between scientists: borrowings, exchanges of technology, and refinements of each other’s devices.

Now, there is another, no less important, methodological question that needs to be addressed. I have a few times presented an extensive list of various versions of the reaction key. To what extent do all the listed examples and modifications of the keys actually constitute one type of devices? Could one refer to the history of one (type of) instrument in this case? It is once again a matter of perspective and of “ontological” choice.

[181] For example, Edwin G. Boring, in his classic “History of Experimental Psychology”, notes that Cattell “*invented* a lip key and a voice key for vocal reactions” (Boring 1929, 523) (italics added).
[182] The modification is discussed by Scripture himself, when presenting his version of the key (Scripture 1893).

Ruth Benschop (1998) poses this same question while dealing with the history of the tachistoscope. As a methodological move, she proposes to approach the problem of defining the tachistoscope drawing on Wittgenstein's notion of family resemblances. This viewpoint allows Benschop to frame the various modifications of the tachistoscope as family members, which share different sets of attributes, and differ from each other by a variety of distinctions. Each particular version of the tachistoscope (or of the reaction key) is thus defined by an interplay of both similarities and differences with the other versions.

I adopt this theoretical stance from Benschop, adding to it a further (pragmatic) argument: grouping different sorts of devices into one family makes it possible to connect disparate digital representations within the archive. Mapping the diverse models and modifications serves as a form of (primary) contextualization of each individual archival representation. Each particular item represents a certain type of the reaction key, which through the type of reaction, or the designer, manufacturer, or even the material, is linked to other items. That is, each individual device acquires meaning in the context of the whole variety of other models. Some lip key displayed in a digital collection no longer seems such an exotic and bizarre device once the entire technological context is in view: biting keys, speech and voice keys, eye lid keys, thumb-and-finger, piano, simple "telegraph" keys.

The resulting panorama of devices reflects the many and varied attempts by turn-of-the-century psychologists to operationalize the subject's response. All of the keys mentioned above provided a solution to one and the same problem: to grasp/ capture/ record various forms of subject reaction, be it finger movement, lip, jaw or chin motion, voice, a runner's move and many others. They were all framed in the same logic: according to a body part whose movement was registered as the subject's response. This panorama of devices explicates some contours of the technological culture[183] of the time, but also materializes some ideas about the subject, their reactions and mental processes. The idea of measuring the mind perhaps here reached its peak, in which the mental processes were evaluated by the movement of a particular organ (at this point my interpretation goes far beyond what ontology can reflect, so I will refer rather to the relevant literature).[184]

183 On the technological culture of experimental psychology, cf. Sokal, Davis, Merzbach 1976, Evans 2000, Schmidgen 2003.

184 Coon 1993, Danziger 1994. On the history of measurement and precision in experimental psychology see Boring 1961, Benschop, Draaima 2000.

3.4.4. Detecting the genealogy of tradition

The biographical sketch that I have charted by consistently registering the constructions of the reaction keys and academic biographies of their creators makes one interesting pattern visible. It reveals that nearly all of the reaction key "inventors" were students (or students of students) of Wilhelm Wundt[185] (Fig. 3.4.4).

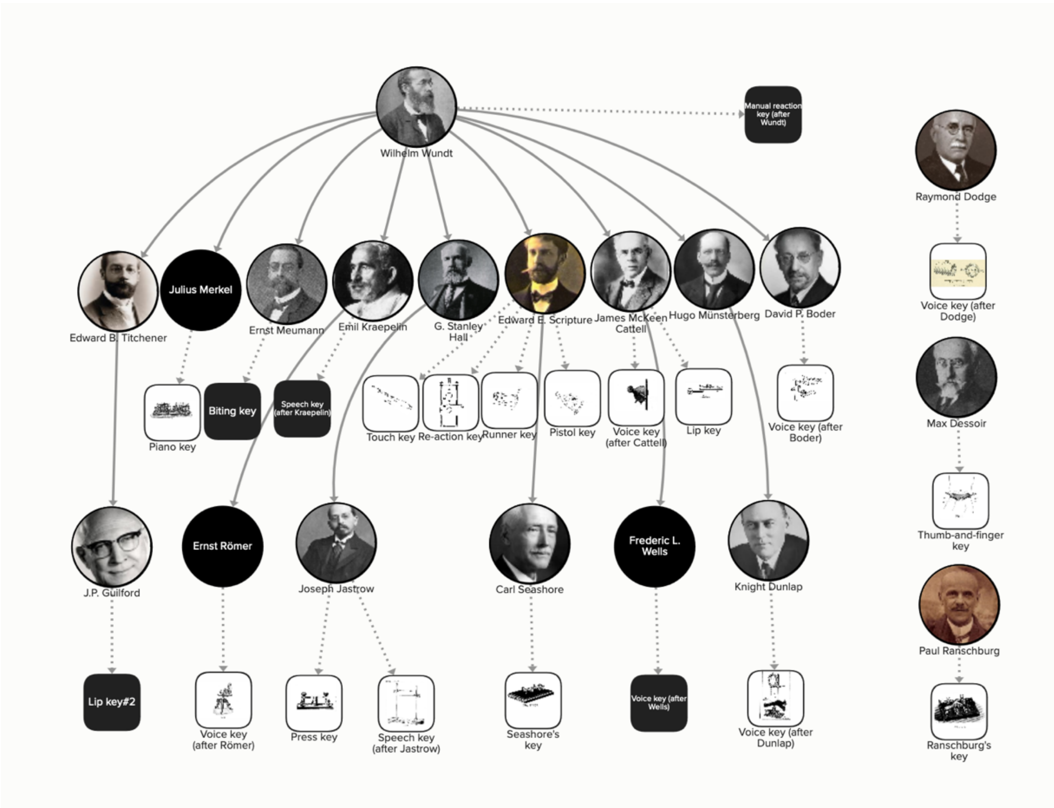


Figure 3.4.4. The reaction keys and their 'inventors'. The solid lines show the professional and academic relationships between people (above all, the relationship between the (doctoral) student and the supervisor); the dotted lines represent the connection between the 'inventor' and his device.

This tradition extends well into the 1930s, a decade after Wundt's death, at a time when the basic tenets of the Wundt school had long since been debunked and challenged[186], and the very paradigm of studying mental process had already been replaced by behaviorism. Still in 1930, J. P. Guilford creates a new version of the lip key (Richards, Guilford 1930), and in 1933, David P. Boder presented yet another modification of the voice key (Boder 1933). Not only Wundt's first-hand students who embraced his system and method, but also the students of the next generation, who often pursued careers in a different paradigm, still, at some point, found themselves constructing new versions of the reaction keys. And even three exceptions to the

[185] It is possible that the way I collected the data or my choice of sources might have influenced my conclusions: I might not have taken some of the keys into account, but even so, the conclusion about the tradition itself remains valid.

[186] E.g., cf. Mülberger 2012.

general pattern were still closely tied into this network of inventions: Dodge's key was a modification of Cattell's voice key, and Dessoir's key, as already mentioned, was modified by Scripture.

The history of the construction of the reaction key family thus forms a unified "academic" tree, with its branches going back to a single root, Wilhelm Wundt. Thus, semantic modelling abruptly makes visible the *genealogy* of the tradition, which is not a matter of transmission of ideas, reciprocal references, or institutional ties. In this case, the very history of designing, constructing, transmitting technology turns out to be genetically linked.

What is particularly interesting here is this inertia (or vitality?) of the technological tradition: it appears to be far more stable than the conceptual, "ideas-based" tradition. Even after Wundt's ideas have been disavowed and discarded, the technological transmission carries on. This pattern becomes visible by virtue of the ontology – the scheme and the successive filling of its points and clauses – but it then needs to be interpreted outside of it. There may be several explanations: I will only quickly list some of them to complete the picture, although they can hardly be modeled within the framework of ontology and would rather call for further purely historical-scientific research.

First, this unity of tradition can be interpreted as a transmission of skills and implicit knowledge that passes from teacher to disciples. Second, one can also discern a social background behind it: the transmission of the view that designing instruments is an integral part of the profession[187]. Third, this pattern can be interpreted not only as a design tradition, but also as a tradition of *communicating* 'inventions': the distinctive ways of presenting, publicizing, naming them. There must have been other reaction keys constructed, but this tradition of 'personalized' keys associated with their creators and presented as inventions, apparently originated with Wilhelm Wundt.

3.4.5. Tracing endings and disappearances

While inventions, designs and modifications of instruments in experimental psychology are thoroughly documented, the events of repair, breakage, discontinuation of use can only be inferred from the *lack* of documentation. That is, fall into disuse, discontinuation, or archiving cannot be modeled as documented *events* – they can only be discerned by vanishing from the

[187] The psychologist-experimentalist, was supposed to be an engineer at the same time, understanding the technical set-up and inventing his own devices. A prime example of such a psychologist-inventor was James McKeen Cattell, inventing the gravity chronometer, lip and voice keys. On Cattell, see Benshop 1998.

landscape, by not being mentioned, by silence. One notable example of such disappearance: the 1930 Stoelting catalog offers no less than ten different keys[188], while the 1937 brochure mentions only one modification (being part of a new experimental set-up of the photopolygraph).

One can also capture the fall into disuse by tracking the dynamics of references to the *usage* of the reaction key. As an example, I traced all the mentions of one modification of the device – the lip key, in English-language publications since 1910 (Fig. 3.4.5).

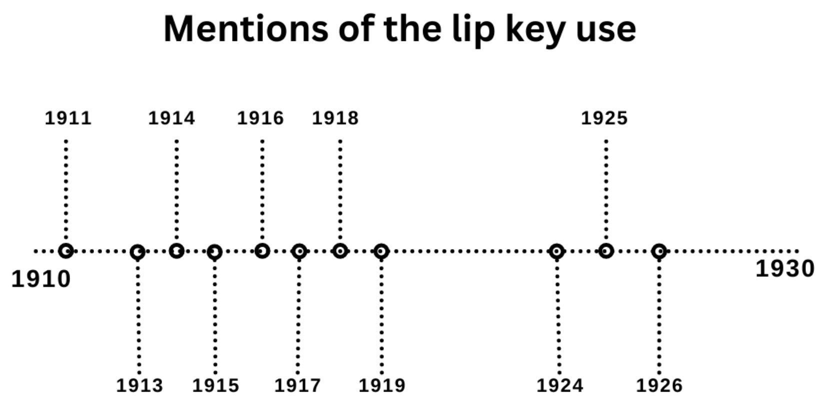


Figure 3.4.5. Mentions of the lip key usage by dates.

The obvious peak of the documented experiments with the lip key falls in the decade from 1910 to 1920; in the next decade three more mentions appear, followed by the last record in 1930. Curiously, however, the 1930 publication does not discuss the use of the lip key in an experiment, but instead presents a new (yet another) modification of the device! So, on the one hand, we can notice a clear decline in the use of the lip key and its progressive withdrawal from use; on the other hand, the modification/ updating of the key goes on (though no information is available about the further fate of this modification).

One can also discern the fall into disuse by tracing down those models of the keys, which remained *in use* outside of experimental psychology, in the 1950-60s. Representations of all these devices are exhibited by digital archives and therefore get into ontology, showing by their very appearance that the history of reaction key continues beyond the experimental psychology of the turn of the century. It is somewhat the mechanics of filling in the points and clauses of ontology, which leads to the conclusion that reaction key passes into the cognitive sciences. In the archives and publications of this period, one again finds the ordinary telegraph

188 Lip and voice keys after Cattell, press and speech keys after Jastrow, biting key, voice keys after Römer and Dunlap, Ranschburg's key, re-action and touch keys after Scripture.

keys, voice keys (and their new modifications), as well as some new modifications, such as response panels or multiple choice apparatuses. Meanwhile, most of the authorial exotic fin-de-siècle models – lip, biting, eyelid, pistol, runner, or touch keys – are no longer mentioned.

3.4.6. Modelling transitions and transmissions

Transitions, appropriations, "adoption of alien objects" (Kopytoff 1986, 67) constitute a focal point of the biographical approach, as formulated by Kopytoff. Through articulating the transitions of things from one context to another, the biographical approach seeks to explore how things are being "culturally redefined and put to use" (Kopytoff 1986, 67) in different social and cultural settings.

In the history of the reaction key one can identify two crucial transitions: first, the (dual) transition from telegraphy to astronomy to experimental psychology; and second, the passage from the turn-of-the-century experimental psychology to the cognitive research of the 1960s. The first case represents a direct adoption/borrowing of technologies from another discipline; the second constitutes a peculiar version of *translatio studii*. The first case is detailed in literature (Schaffer 1988, Canales 2001, Schmidgen 2003) as part of the well-elaborated history of the Wundt's set-up for the reaction experiment; the second is barely known through some accounts of the technological development in psychological research[189].

For modelling in ontology, such transitions pose a challenge. One problem comes from the fact that the transition constitutes not a single event, but a group of events comprising many different micro-transactions: conversations, citations, translations, encounters, borrowings. Transition, on the other hand, is more of a historical assertion, an a posteriori synthetic conclusion, only possible once it has already occurred. Another problem is that the transition refers not to a concrete thing which can be specified within the ontology, but rather to some (technological) knowledge in transit.

Still, the transitions seem to be too valuable for the contextualization of the artifact to be overlooked. Take, for example, the following three archival representations (Fig. 3.4.6.):

[189] Some continuity between the two traditions is noticed in McPherson 1979, Caudle 1983, Robinson 2001.

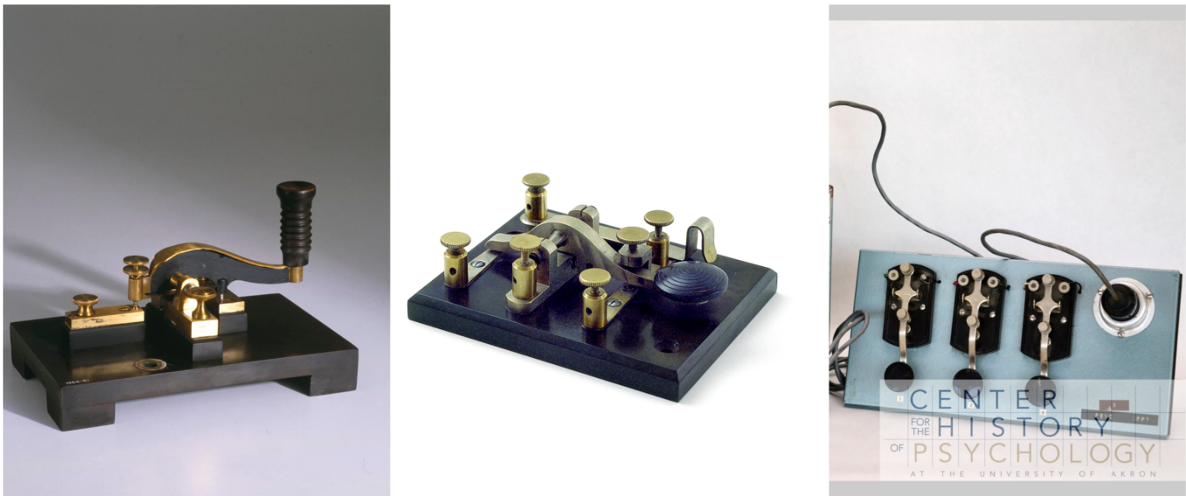


Figure 3.4.6. Three 'telegraph' keys from digital collections

The first is a telegraph key used to transmit messages in the 1850-70s[190], the second is a manual reaction key used in experimental psychology at the turn of the century[191], and the third is a number of connected telegraph keys used in psychological experiments, presumably after the 1950s[192]. How can we show, on the one hand, the continuity and connection between these three instruments, and, on the other hand, the fact that they operated in completely different research contexts?

My solution was to trace the *micro-events* of transmission which 1) are documented (the ontology specifies the documentation), 2) can be associated with an entity defined within the ontology (e.g., the type of the reaction key, particular artifact or particular concept). To this end, I have established a special subclass of the 'event' – 'transmission', which includes various kinds of personal communication, textual transmissions and technological adoptions. Consequently, each great transition in the ontology appears as a set of such events. So, the transition of the telegraph key to the context of experimental psychology is modeled as follows:

> 1) Adolph Hirsch adopts telegraph key for his experiment (Hirsch 1864, fact of technology adoption) -> 2) Wilhelm Wundt adopts Hirsch's set-up, and references his publication in his 1874 edition of "Grundzüge der physiologischen Psychologie" (fact of borrowing and citation) -> 3) Carl Krille, a technician at the Leipzig laboratory, modifies the telegraph key for Wundt's needs.

---

190 Science Museum Group, [https://collection.sciencemuseumgroup.org.uk/objects/co33379/morse-key-1850-1870-telegraph-peripheral]

191 This is one of the modifications presented in the Zimmermann catalogs. Exhibited by the University of Belgrade (The collection of the old scientific instruments, Laboratory for experimental psychology). [http://lep.rs/en/about-us/the-collection-of-old-scientific-instruments/miscellanea/reaction-key-e-zimmermann-leipzig-berlin-2/]

192 The Archives of the History of American Psychology, Cummings Center for the History of Psychology, the University of Akron. [https://collections.uakron.edu/digital/collection/p15960coll7/id/436/rec/8]

In this way, through a successive chain of transmission events, the telegraph key technology passes from telegraphy to astronomy to experimental psychology. This sequence of events appears as a pretext for a further surge of its 'inventions' and modifications. It indicates the origins of the key as a psychological device by linking the tradition of experimental psychology to both astronomy and telegraphy.

The second transition of the key is still more intriguing. In this case we are dealing perhaps with a less radical disciplinary divide (astronomy/psychology vs. experimental psychology/cognitive psychology), yet with a greater time gap. The period of the 1920-50s is known in the history of psychology as the era of Behaviorism, "the dark age of mental chronometry" (Meyer et al. 1988, 11-12) when the paradigm of the reaction time research was pushed to the periphery. In the 1950s and 1960s, when the discipline of cognitive psychology began to take shape, one again found in experiments the telegraph keys, voice keys, and eventually the "successors" of the fin-de-siècle devices: switches, response panels, multiple choice apparatuses. What is at stake here in terms of modelling the digital archive is showing the continuity of different types of devices and situating each of them in the context of its successors and predecessors.

To begin with, the ontology makes visible that the tradition was not disrupted: the reaction keys continued to be designed in the 1930s and 1940s and apparently continued to be used. The consistent and meticulous tracking of events in the history of a thing, as required by ontology, thus elaborates and complicates the classical version of the history of psychology, written rather from the standpoint of the history of ideas. It reveals that in the era "from Watson (1913) to Skinner (1963)", when "nothing seemed to be happening" [193] in the study of cognitive processes, an instrumental (and so apparently experimental) culture of mental chronometry continued to operate. Moreover, not only does it show the very fact that the technology continued to evolve, but it also traces the continuity between early and later devices.

One of the very few studies comparing instruments in experimental and cognitive psychology is Fairfid M. Caudle's 1983 article, which traces "a surprising degree of continuity" between the two disciplines (Caudle1983, 21). Caudle's method is as follows: he demonstrates one example of a turn-of-the-century instrument and one example from the "modern" (1980s) psychological laboratory, commenting on their similarities and differences. For instance, he shows an "early version of the voice key" (probably Cattell's model) and then a "modern"

---

[193] As described in Ulric Niesser's "Cognitive psychology" (1967, 5). This same "dark period" of Behaviorism is traced through the *American Journal of Psychology* publications in O'Shea, Bashore 2012.

version produced by Lafayette Instrument Company, observing that "the ugly duckling of the early voice key has developed into a sleek swan, the sensitive and much-streamlined electronic voice reaction time apparatus" (ibid., 26). The ontology elaborates on this account, by making it possible to trace a multitude of intermediary steps between these two poles[194]. In the voice keys example, it is able to show the first model with a microphone added (Boder 1933) or the one in which thermionic relays come into use (Kahn 1935). Not only these purely technical adjustments become noticeable, but also the relation of the device to the wider (technological) culture. In early experimental psychology, and particularly in Wundt's laboratory, the instruments were adopted from the physics laboratory and communication technologies (as highlighted by Schmidgen 2003). In the 1920s and 1930s, however, the designers of the voice keys found inspiration in the artifacts of popular culture. So, Frederic Lyman Wells, a psychologist from Boston Psychopathic Hospital who presented a new device in 1922, took inspiration from the two-dollar "Radio Rex" toy (Wells, Rooney 1922), and a decade later David P. Boder took its cue from a toy called "Microphone dancer" (Boder 1933).

Figure 3.4.7. "Radio Rex" and "Microphone dancer"

Such an appeal to popular culture would hardly have been possible in the Leipzig laboratory, with its cult of metrology, discipline, and its claims to produce "true science" (Benschop, Draaisma 2000). To give a further example, the devices of the 1970s and 1980s appeared to be much more closely intertwined with media technologies: first came the practice of converting

194 The voice key example is quite well covered in the corresponding section of the Key2Mind exhibition: https://key2mind.omeka.net/exhibits/show/biography/overview/vocal

tape recorders into voice keys[195], and then the reaction time software[196] appeared, which was inspired by the aesthetics of video games.

Apart from technological succession, the transition of the key to cognitive psychology involved various forms of textual transmission. In this case the chain of knowledge transfer was longer and less linear than in the passage from astronomy to psychology. Tracing this transmission properly would require a separate study, so I will cite but one example: the 1952 article by the British psychologist William Edmund Hick, entitled “On the rate of gain of information”. In this article, Hick presents the multiple-choice experiments that led him to formulate what is known as Hick’s law: a logarithmic relation between the decision time and the number of choices.

The experiment conducted by Hick echoed the work of Julius Merkel, one of Wundt’s doctoral students, who in his doctoral dissertation (Merkel 1885) investigated *Wahlenzeit,* the time of volition or the will-time[197]. Unlike Wundt’s other doctoral students, Merkel apparently did not become an eminent psychologist and a laboratory founder – in fact, apart from studying under Wundt, I was unable to find any information about him. Hick comments on Merkel’s experiments as follows:

> Merkel himself was chiefly interested in the supposed divisibility of reaction time into “cognition time” and “choice time,” and does not even give the raw data. However, they are tabulated by Woodworth (1938)… (The original paper is not very accessible, and the writer is indebted to Mr. A. Leonard for obtaining it and translating the relevant parts.)… An interesting piece of experimental technique is the use of a Geissler tube in order to ensure the sudden onset of the illumination. The Geissler tube was an early form of gas-discharge tube, the forerunner of the modern fluorescent lighting. The illumination would be rather weak, but doubtless quite adequate if the subject was moderately dark-adapted. Some of the other archaisms, however, are less pleasing. For example, there is no indication of the order in which the stimuli were given… (Hick 1952, 11).

The first thing worth looking at in this quote is the very complexity and multi-stage nature of the transmission. Hick had to rely on someone’s translation of Merkel’s work and to draw the data not from the original, but from the influential textbook by Woodworth, which proves to be an important mediator between the two traditions. Robert S. Woodworth himself was a doctoral student of James McKeen Cattell, one of the first doctoral students of Wundt. His textbook, written in 1938, has been revised and reprinted several times; in particular, the 1954 version, co-authored with Harold Schlosberg, was highly influential and cited. Thus, one

---

[195] Some examples to follow: Wilkinson, Houghton 1975, Perera 1980.

[196] Some examples of the 1980s software are stored in the Internet Archive: [https://archive.org/details/d64_Reaction_Time_1981_Andrew_Colin], [https://archive.org/details/d64_Reaction-Time_Tester_1985_Tab_Books]

[197] The “will-time” is J. M. Cattell’s translation of *Wahlenzeit* (Cattell 1886c).

discovers an entire succession of knowledge transmissions that stretches all the way back to Wundt himself.

Hick's commentary on Merkel's experiments is also noteworthy. Hick speaks of "archaisms," pointing to the "obsolete" practices for describing the experiment (e.g., no description of the order of stimuli given), the "obsolete" problem statement ("*supposed* divisibility of reaction time"), and the "obsolete" technology (e.g., his comment on the Geissler tube). Yet, substituting all three listed components, Hick essentially repeats Merkel's experiment: he even integrates Merkel's findings into his own data! Furthermore, like Merkel, he makes use of several connected telegraph keys, a device which in experimental psychology was called a "piano key" (Fig. 3.4.8).

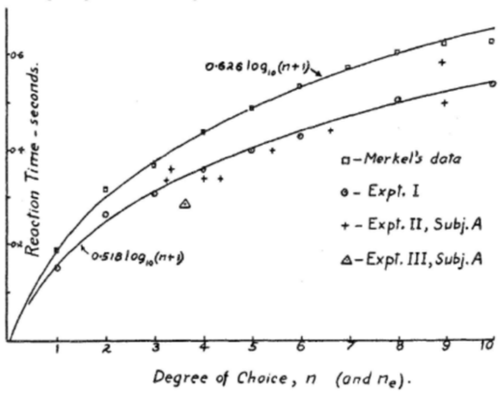


Figure 3.4.8. On the left is William Edmund Hick with his instruments for the multiple-choice experiment, including ten connected telegraph keys (from Reynolds, 2004, 90).
On the right, is one of Hick's graphs in which he merges Merkel's and his own data (Hick, 1952, 15)

Hick, in turn, also appears as an important mediator: by conducting his experiments a decade before cognitive psychology was formed as a discipline, he established a reference point[198] for the numerous multiple-choice experiments that would later be conducted. Subsequent choice experiments are descended from Hick and via him, in turn, from Merkel.

3.4.7. Contrasting research contexts

Concluding the biographical sketch, I will briefly comment on the two research contexts or paradigms, in which the reaction key finds its use and meaning: experimental and cognitive psychology.

The very *research context* class is quite problematic in the ontology. In the reaction key example, it has actually come to represent a context of a discipline: in this way, I modelled the contexts of *experimental psychology*, referring to the studies of the 1880s-1900s and *cognitive psychology*, referring to the paradigm of the 1960s and 1970s. Yet, obviously, the boundaries of any discipline cannot be drawn clearly: for example, it is not always possible to make such a sharp distinction between *experimental* and *applied* psychology, or to determine the starting point of cognitive psychology as a discipline. Any labeling of this kind is thus inevitably contingent. Not to mention the fact that strictly speaking, cognitive psychology is also experimental.

Another issue is that such broad categories, as *experimental* or *cognitive* psychology**,** obscure the many differences within these "paradigms": following the ontology, one cannot distinguish, for example, between the tradition of the Leipzig Laboratory and that of the Würzburg Laboratory, or separate Ulrich Neisser's "humanistic" approach in cognitive psychology from Michael Posner's more "rigorous" method.

That said, these two roughly sketched research contexts make it possible to draw some comparisons between the two traditions and, through contrast, to get a sense of their peculiarities. The "personalistic" tradition of the reaction keys in experimental psychology becomes apparent once one adds to the points and clauses of the ontology the 1950-70s devices, which happen to be anonymous and impersonal. In the ontology, I initially adapted the pattern of experimental psychology to identify the keys (referring to "Cattell's lip key" or "Boder's voice key"), without even noticing it, and only when I reached the 1950s, did I realize that the *pattern* had changed.

Keys in "cognitive psychology" can no longer be traced in the same way as those in "experimental psychology": they have no "names," "creators" or representations, they are

[198] On the influence of Hick's article, see (Proctor, Schneider 2018, 1282). Based on citation indices, the article continues to be regularly cited (according to the Web of Science, about 30-40 citations per year between 2006 and 2016).

neither "invented" nor documented, and paradoxically, they are much harder to find in the archives (at least in the digital ones)[199]. In cognitive psychology, the source of interest, "innovation" or "discovery," never comes from the device *per se*[200]: all the attention is given to the interpretation of the experiment, the data, graphs, and their discussion. The reaction keys appear in glimpses and are referred to as types rather than individual modifications: as a rule, there is only a brief mention, for example, of the usage of some "electronic voice key" in an experiment, without specifying the model or the manufacturer.

This configuration brings to light the role of the devices back in experimental psychology, where the reaction keys were individualized and named after their prominent "inventors"; where each new modification was presented, published and talked about; where the reaction keys appeared as part of a shared framework of a thought collective (*Denkkollektiv*) of experimental psychologists.

[199] Somehow even the catalogs of psychological instruments after the 1940s have not been digitized.
[200] The only exception to this rule that I am aware of is psychologist Arthur Jensen, who in the 1980s 'invented' his 'signature' reaction device called the Jensen box (Jensen 1982).

3.5. Assemblages

3.5.1. Theories of assemblages

Whereas the biographical model situates an object within a diachronic perspective, this second model is meant to get closer to capturing *science-in-the-making*. The assemblage gives rather a synchronic snapshot of how and in what circumstances the artifact acted and received its meaning; with what instruments, people, representations, it came into interaction and 'co-operated' in one setting or another.

The notion of assemblage offers a way of reflecting and conceptualizing these forms of co-operation, arrangement, ordering, or distribution of heterogeneous agents. Assemblage thinking assumes that entities take shape and meaning in the process of interaction with other entities. That is, all entities are relational: "they are produced in relations"[201], in co-operation, in concerted action, alliances and combinations. As Latour succinctly puts it, "attachments are first, actors are second" (Latour 2007, 217) This common understanding is present in the various conceptualizations of assemblage, from the original definition of *agencement* by Gilles Deleuze and Félix Guattari to Actor-Network theory (Law, Hassard 1999, Latour 2007) and the assemblage theory of Manuel DeLanda (2016).

Describing an object through its relations with other objects, this model is designed as an alternative to the archival practices of situating things. The archive conserves, isolates and singularizes its objects. In it, a thing is performed as a unique, self-contained and stand-alone artifact, removed and detached from its everyday settings and pragmatic context. As discussed in Part II, even in those (rare) cases where the archive connects objects, these relations tend to be rigid, restrictive, similarity-based. It is this isolation that this model is designed to overcome by situating objects in the context of the assemblages, material and social configurations in which they have engaged and interplayed with other agents.

In this sense, I draw on Actor-network theory or the perspective of "material semiotics" (Law 2008), which frames laboratory settings as "assemblies of humans and nonhuman actants where the competences and performances are distributed" (Akrich, Latour 1994, 259). The experimental set-up appears as a certain configuration made up of humans and 'non-humans': machines, people, concepts, inscriptions, theories, phenomena, materials, which act together.

---

201 As John Law puts it, "It [ANT] takes the semiotic insight, that of the relationality of entities, the notion that they are produced in relations, and applies this ruthlessly to all materials— and not simply to those that are linguistic" (Law, Hassard 1999, 4).

Within a given configuration, each of the participants plays out certain scenarios (or *scripts*)[202]. Thematizing the assemblage as a constellation of heterogeneous agents (or *actors* and *actants*), this perspective challenges the strict divisions between things of nature and things of culture[203], as well as between human and non-human agents.

Yet another key concept I should elucidate before moving on to the model is the notion of inscriptions. Inscriptions are "all traces, spots, points, histograms, recorded numbers, spectra, peaks, and so on" (Latour, Woolgar 1986, 88), these are all marks or signs recorded (*inscribed*) in the course of scientific practice, what is "behind a scientific text" and what is obtained by "setting up instruments" (Latour 1988, 67). While the importance of inscriptions as immutable mobiles has been discussed in the previous chapter, in the context of the assemblage model, inscriptions are important insofar as they are produced in the course of the "experimental assemblage" and take part in it. By "recreating" some fragments of the relationships between humans, instruments, the inscriptions they produced and the concepts next to which they were used, the model is meant to recreate the pragmatic context of an artifact within which it operated and gained meaning.

3.5.2. Modelling assemblages

The model articulates interactions of an object with humans and "non-humans", describing the objects as being part of assemblages, constellations and networks. It is designed to meet the following questions: In what settings has the object acted? What configurations gave the object its function, meaning and identity?[204] What artifacts, people, concepts, theories, inscriptions has it interacted with? How has its role (or script) changed over time?

The model is structured around *assemblages*, in which different types of *agents* take part and perform *actions* on each other. The core class of the model is the *assemblage*, bringing together human and non-human agents. In the ontology, *assemblage* can define, for example, an experimental set-up, which *includes* the instruments operated within it, its experimenter and subject, the inscriptions made during the experiment, the conceptual objects feeding the experiment (Fig. 3.5.1).

---

[202] The Actor-network theory has an entire vocabulary to describe how these scripts can change: translation, delegation, shifting in and shifting out, displacement (Latour 1994; 1995).

[203] On the principle of symmetry between nature and culture, human and non-human, see Latour 1993.

[204] ANT and assemblage theory propose that relations *precede* the object – a premise that cannot be reflected in a semantic model, which supposes the preexistence of classes of things.

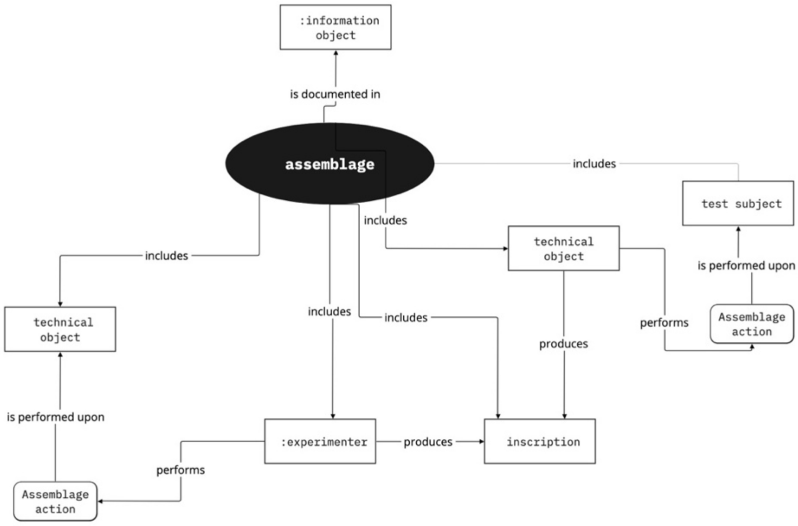


Figure 3.5.1. Schematic view of the model (fragment)

Within the assemblage, agents can perform actions, engage in interactions with each other. To keep track of these actions, I introduced a distinct *Assemblage action* class (such that each action necessarily has an assemblage context). Within the ontology, *assemblage actions* are *performed by* an agent and *performed upon* another agent in the context of the *assemblage*.

Further, the *assemblage* can be related to the class of *conceptual objects* by a number of different connections: it may *have a research question* and may *lead to/ result in* a *theory* (*research question* and *theory* are modeled as subclasses of the *conceptual object* class). An assemblage can also be situated in a *research context*. In addition, the model makes it possible to trace connections between assemblages: so, an assemblage can *inform* or *be informed* by another assemblage (Fig. 3.5.2.).

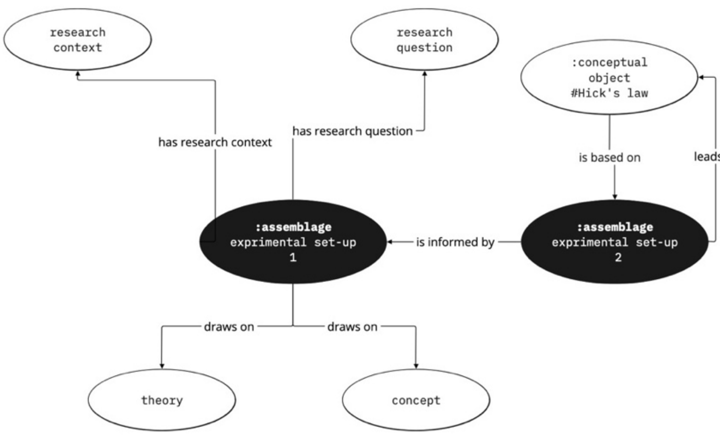


Figure 3.5.2. An example of the contextualization of assemblage in the model. Assemblage not only relies on concepts and theories, but can also lead to some new concepts and theories: as in the case of Hick's law.

3.5.3. Weaving things together through assemblages

In order to show the model in action, I modeled eleven assemblages – experimental set-ups featuring the reaction keys[205]. In this case, I focused primarily on the set-ups of experimental psychology and sought to sample very diverse examples: experiments from across the years, employing different models of the reaction keys, and covering various research topics. To illustrate the modelling of the relations between assemblages, I also included two more recent set-ups, rather related to early cognitive psychology, which explicitly refer to two earlier experiments.

My main source in this case was publications: for each of the assemblages, I traced human and non-human agents involved in the experiment, including inscriptions, relevant concepts, and research questions. Publications as sources have a number of limitations: mainly that they *represent* the experiment according to certain conventions. So many of the questions posed by Actor-network theory and many of the conclusions made through direct observation of the experiment cannot be made[206].

The *Assemblage* model yields less material for a historical narrative (and even for this particular text here) than the biographical model. There are no (narratable) events but only (socio-)material configurations, which often resist narration, being not easily expressed in language. Yet it is capable of contextualizing objects beyond language, through its juxtaposition with other agents (in the archive, primarily with other objects it has interacted with). This tying things together is different from cataloging or indexing objects. This can be clearly seen when comparing the virtual Museum of the History of Psychological Experimentation[207] with any assemblage from the *Key2Mind* exhibition (Fig. 3.5.3). The museum, which follows in detail the 1903 Zimmermann catalog, gives a very detailed picture of the material equipment of psychological laboratories at the turn of the century, but does not associate the instruments with one another. Assemblage does not presuppose a description of all that is available, but seeks to meaningfully arrange, situate things, to put the thing back into

---

[205] All the eleven set-ups are reflected in the corresponding exhibition: [https://key2mind.omeka.net/exhibits/show/assemblage/set-ups]. It is noteworthy, however, that the interface of the Omeka platform severely impairs the representation of the assemblage. By default, the exhibited items and the text reside in distinct areas of the screen. This brings, for example, the experimenter and the instruments (visually) separated from each other, which is certainly not in the spirit of the very concept of assemblage.

[206] Nor have I had a case where inscriptions made in the course of experiments (rather than inscriptions found in publications) would have been available in the archives.

[207] The museum was created by Edward J. Haupt and Thomas B. Perera at Montclair State University in the late 1990s. It is still available at [http://tomperera.com/psychology_museum/museum.htm]. For more details on the idea of the museum, see Haupt 1998.

its context of action, into a dialogue with other things. In this way, each of the things is contextualized through its relation to other things, as well as (ideally) to human agents, inscriptions, concepts.

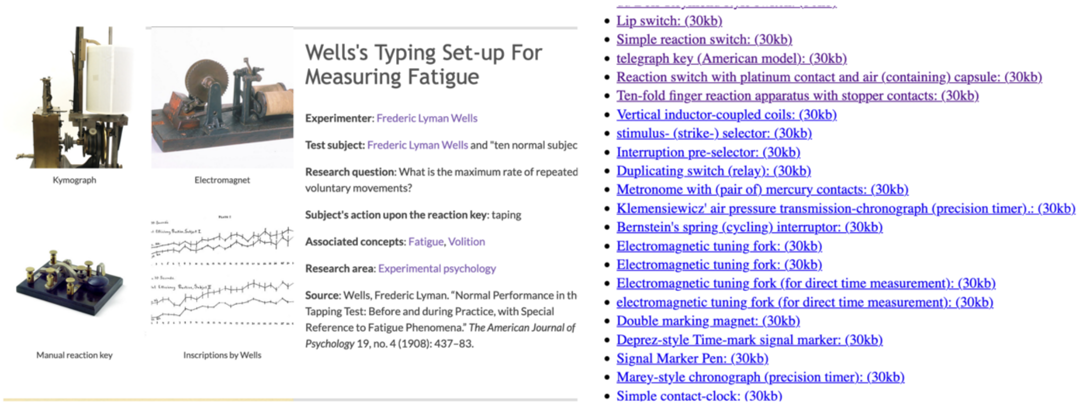


Figure 3.5.3. Examples of the assemblage at the Key2Mind exhibition (left) and of the instruments arrangements in the Museum of the History of Psychological Experimentation (right).

Beyond the encounter of things in the archive, the model has the potential to elaborate and animate existing historical accounts – in our case, those dealing with the material culture of experimental psychology. The studies addressing the material arrangements of early psychological laboratories tend to focus on one particular instrument – the Hipp chronoscope, a device for measuring time to the nearest one thousandth of a second. To use Horst Gundlach's words, the Hipp chronoscope is a "totem pole" of the experimental and applied psychology tribe (Gundlach 1996). Adopted by Wundt from astronomy (along with the key), the Hipp chronoscope became an indispensable part of the reaction experiment (for good reason its name appears in the title of the famous Wundt's set-up) and made its way into all major psychological laboratories.

In historical research, the chronoscope is interpreted as an essential part of the material arrangement and laboratory work of late nineteenth century psychology, but also as a versatile metaphor for the entire paradigm[208]. It is viewed as an instrument embodying and symbolically expressing the ideals of precision peculiar to the culture of early experimental psychology (Benshop, Draaisma 2000, Schmidgen 2003 and 2005).

Yet, as Gundlach adds, the chronoscope was always accompanied by the "indispensable accessories" that made the reaction experiment possible. Among these "minor items" (Gundlach 1996, 68) he lists the rheochord, the rheostat, galvanic elements and the reaction

[208] For example, Benshop and Draaisma (2000) describe Wundt's research program through the metaphor of mind *calibration*. Cf. also Canales 2009.

keys. Electrometers, commutators, electromagnets can also be added to the list. It was these minor elements that underwent modifications as the set-up became more complex, while the chronoscope itself remained virtually unchanged[209]. If at all mentioned in the studies of early psychological laboratories, these elements are usually referred to in passing, as a quick enumeration, while the chronoscope is given a major technological and symbolic role.

The assemblage model offers a way to deconstruct these hierarchies and *horizontalize* the components of the assemblage. Consistently and meticulously mapping out all the galvanometers, magnets, and Daniel cells, the model articulates the place of these "accessories" in various set-ups without distinguishing between major and minor devices. Within the archive, it becomes then possible to visualize, animate, make manifest the multiplicity of instruments within a single experimental setting. Such mappings could bring attention to the extent that early psychological set-ups adopted instruments from physical laboratories and communications research[210]. Or they could give evidence that conducting experiments involving all this equipment required some expertise in electrical circuits. Finally, they make it apparent that not only conceptually, but also materially, the early psychological set-up was embedded in the technological and scientific culture of the turn of the century.

3.5.4. Contextualizing the reaction key within the assemblage

Before turning to the various forms of distant reading, I will zoom in on one particular assemblage in order to examine how it contextualizes the agents. Consider, as an example, the simplest set-up for which I modeled all the interactions of the agents or "assemblage actions" (Fig. 3.5.4).

The schema represents a classic Cattell experiment with the gravity chronoscope, the Hipp chronoscope, and the lip key, used to measure the time of naming objects, words, and colors (Cattell 1886c)[211]. On command, the experimenter triggers the gravity chronometer: the screen drops, so that the subject can see a stimulus (e.g., a word written on a card). At the same moment, the falling screen sets the Hipp chronoscope in motion. Once the subject reads the

209 This is noted by Gundlach himself (1996, 68). The same tendency is observed by Ruth Benshop and Douwe Draaisma, when comparing the set-ups of the reaction experiment in two different editions of Wundt's "Grundzüge der physiologischen Psychologie" (Benshop, Draaisma 2000, 13).

210 All of these accessories are not specifically psychological instruments. Perhaps that is why they have hardly survived in psychological collections, especially compared to the widespread presence of the Hipp chronoscope. Most of the examples I cite in the exhibition come from physical and industrial collections, not psychological ones.

211 The experiment is further detailed in Sokal, Davis, Merzbach 1976 and Benshop, Draaisma 2000, 15-19. See also the representation at the Key2Mind exhibition [https://key2mind.omeka.net/exhibits/show/assemblage/set-ups#exp3].

word, the lip key, which responds to the movement of his lips, stops the chronoscope. The chronoscope shows the time of the response, which is recorded down to become an inscription.

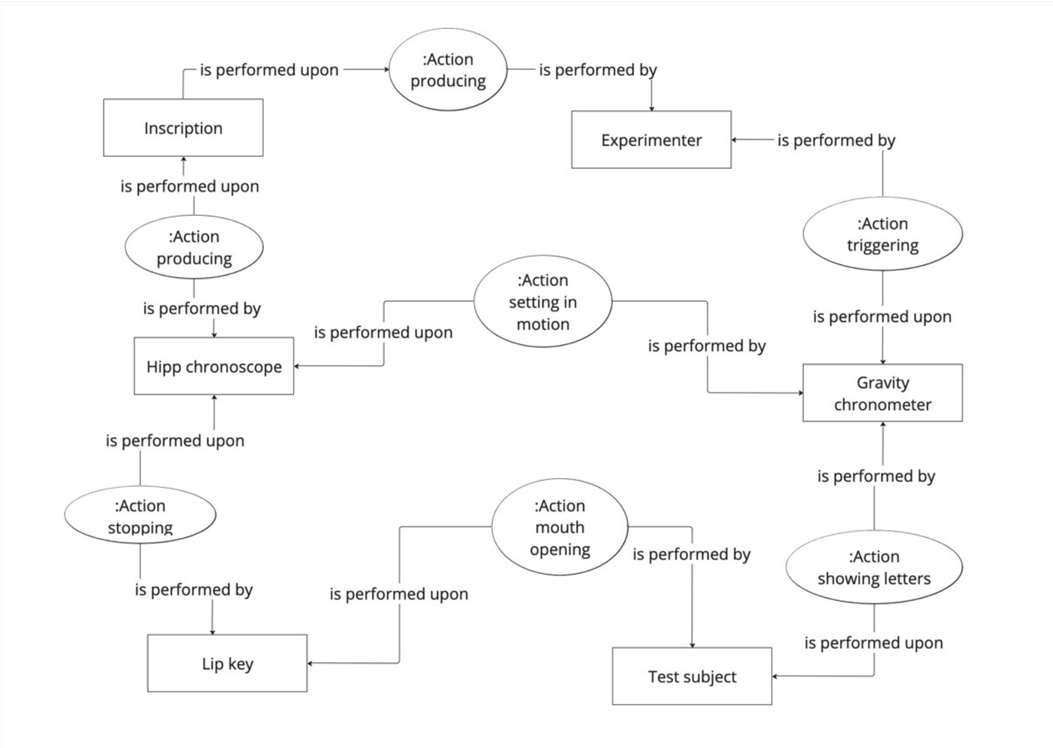


Figure 3.5.4. Schematic representation of J.McK. Cattell's naming experiment.

The schema represents an almost automatic domino-like chain, in which every agent triggers, sets in motion or stops every subsequent agent. In this chain of triggering and stopping, all agents take an equal part and have essentially the same function – to set in motion or stop each other. This applies not only to the instruments, but also to the experimenter and even to the subject acting on an equal footing with the non-human agents. As a result, this "automatic circuit" generates a digit – a reaction time to the nearest one thousandth of a second. This digit is written down; a number of those digits are then put together into graphs and form the basis for statistics[212]. The assemblage actions and interactions of agents would probably be challenging to model in the archive, yet the context of the inscriptions produced by a particular instrument seems to be particularly vital to the contextualization of scientific residues. It allows outlining a certain horizon of the instrument use and action, to show what kinds of facts it gave birth to. In the case of the reaction key, these are mostly numbers that are meant to serve as answers to the questions experimental psychology asks.

212 On the complex relationship between experimental psychology and statistics, see Nuttgens 2023.

To further contextualize this material configuration for producing “matter of facts”, I also associated the assemblages with the research questions they were meant to answer and with the concepts (in ontology, *conceptual agents*) on which they relied (Fig.3.5.5). This further contextualization brings to light the most exciting part: the fact that this whole material and technological setting, statistics and graphs were meant to answer questions about the mind, the will, impressions and apperceptions. What has traditionally been considered immeasurable philosophical concepts, in experimental psychology turns out to be the subject of measurement. For example, the naming experiment draws on the notion of impression, formulated in David Hume’s philosophy and understood in psychology as an immediate experience of encountering an object (for example, a stimulus).

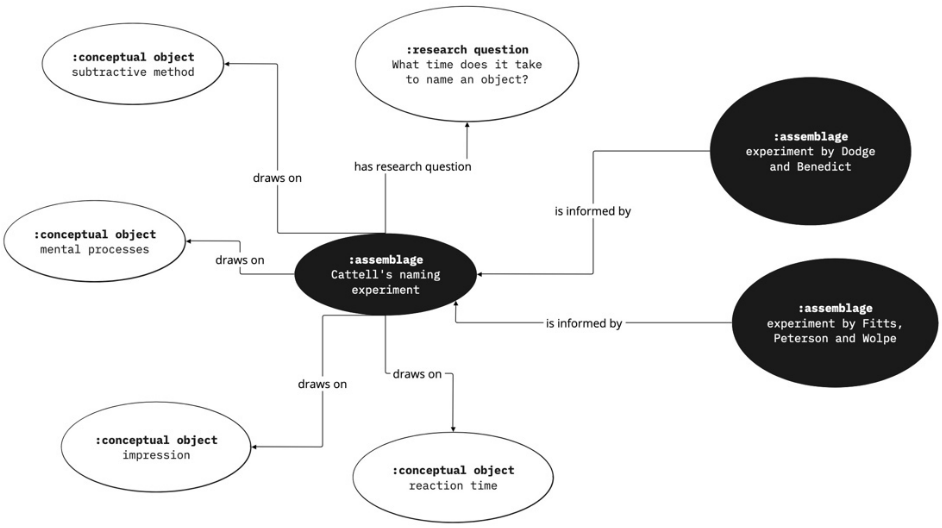


Figure 3.5.5. Further contextualization of the Cattell’s experimental set-up

Beyond that, the assemblage also operates in context: first, in the context of a discipline (modeled in ontology as “research context”), and second, in the context of other assemblages. Experimental set-ups replicate, adopt, elaborate, modify, are inspired by or contest with other set-ups and experiments. These relationships can also be reflected in the ontology.

In this way, each assemblage is also placed in the context of other assemblages: as their source, or as their sequel. So, for example, Cattell’s naming experiment informs two other experiments: Dodge and Benedict’s experiment to measure the effects of alcohol on word-reaction (Dodge, Benedict 1915) and late set-up by Fitts and others to measure the correlation of reaction time and stimuli frequency (Fitts, Peterson, Wolpe 1963).

The assemblage model thus entails a multistage or multilayered contextualization of an object. The assemblage first forms the socio-material context in which various agents encounter and act together, and then it is itself situated in the framework of the discipline and other assemblages, thus forming *an assemblage of assemblages*.

3.5.5. Reading across assemblages

The major power of ontology, in fact, manifests itself not in close-reading as discussed above, but in making possible the *distant reading* of assemblages. The ontology offers the possibility to track assemblages and their various agents in development. It allows one to trace how a particular agent's relationships evolve, with which agents it meets more often than the others, how the technological environment and practices of inscriptions change[213]. I will give some examples of what conclusions can be drawn from such an analysis (although my examples will be limited to a small sample of eleven set-ups).

In the first place, the ontology offers a statistical view of the technological context: i.e., the frequency of encounters of the instruments with one another across different assemblages. As advocated in the part II, such data provides a meaningful ground for the production of links within the archive: it puts an artifact within the (original) context of things, in which it operated and gained meaning. Such statistics would hardly be meaningful on the basis of only eleven assemblages, so at this point I will limit myself to a few very general remarks. To begin with, even according to my (quite negligible) sample, the reaction keys are most often found together with the Hipp chronoscope. One can also trace the Hipp chronoscope going out of use and being replaced by other instruments for measuring time (notably the Jaquet chronograph in the 1915 experiment by Dodge and Benedict (1915), and the simple 0.1 sec clock in the 1960s set-ups). In contrast to the chronoscope, which fell out of use at some point between experimental and cognitive psychology, the stimulus device shows some permanence: the 1960s experiments still make use of the tachistoscope, as did Cattell's 1886 and Dodge's 1915 experiments.

Further, the assemblage model can provide a glimpse into the ways in which human agents are engaged in the experiments. The literature has already detailed the practices of switching roles between experimenter and subject in early experimental psychology[214]. The

[213] These capabilities of ontology are also hardly reflected in the exhibition. The Omeka interface does not allow combinations and re-combinations of exhibits, i.e. one cannot make queries such as: which concepts operate alongside the lip key? or in which assemblies does the lip key take part together with Hipp chronoscope?

[214] See, for example, Benshop, Draaisma 2000, 19. On social organization in experimental psychology, see Ash 1980, Danziger 1994.

model shows this tendency quite clearly[215]: in the early set-ups, the experimenter also takes the position of the subject. Interestingly, for the first time (in our sample) the roles of experimenter and subject get strictly separated in an experiment from applied psychology, the experiment of Dodge and Benedict (1915) on measuring the effect of alcohol on mental processes. One cannot draw conclusions from a single example, but it seems quite plausible that the demarcation of the two positions takes place precisely in applied and industrial psychology, when theoretical questions are replaced by practical performance measurements (Rabinbach 1992), and the very number of subjects within a single experiment reaches industrial scales. Interestingly, in Hick's experiment conducted in the 1950s, the same pattern appears once again: Hick himself acts as both experimenter and subject.

Another interesting tendency about the human agents concerns the practices of representation of the subject in experiments. In the ontology I have recorded how subjects are characterized in publications, giving exact quotations, for example: "15 graduate and medical students" (Brown 1937) or "8 normal and 3 psychopatic subjects" (Dodge, Benedict 1915). This makes it possible to trace how these representational practices shift from one set-up to another. In early experimental psychology, subjects are usually called by name (and often by profession). Personification of the subject is thought of as one of the means to verify the experiment (and corresponds, I believe, to the personalization of the reaction keys). By the beginning of the century, there appear impersonal descriptions fixing the number of subjects and their mental status: e.g., "ten normal subjects" (Wells 1908) (this practice should probably also be associated with the rise of applied and industrial psychology). In the 1960's, our only example shows an indication of the form of payment (or motivation for the subjects' participation in the experiment): "48 male college students, paid by the hour" (Fitts et al. 1963). Again, the small sample does not allow us to draw conclusions. However, it seems that a quantitative study of these subject-naming practices could uncover a lot of insights into the discourse about the subject within experimental and cognitive psychology.

Aside from the instruments and human agents, the conceptual context of assemblages could also be examined through quantitative analysis. The model allows conducting a search for different set-ups sharing the same concepts, or, conversely, for similar (or related) set-ups described by different concepts. For instance, one can trace the experiments that are associated

[215] With a larger sample of set-ups, it would be possible to construct a network of experimenters and their subjects; that would make sense for a study of a small collective or scientific community, such as the Wundt Laboratory of the 1880s-90s.

with the concept of volition[216]: Merkel's set-up for measuring the time of volition (Merkel 1885), Scripture's set-up for measuring volition (Scripture, Moore 1893), and typing set-up for measuring fatigue by Wells (1908). Volition – understood as the acts of will, motivational activity, such as acts of choice or decision – constituted one of the central concepts of Wundt's system (which even gave its name to the entire system, *voluntarism*). The three experiments dealing with volition measure, adopt, or as we would say today, operationalize this notion in a different way. Merkel, for example, measures volition time in a disjunctive (or multiple choice) experiment, Scripture, and Wells draws on this notion to describe the nature of the subject's movements in his typing test for measuring fatigue.

Another (and reverse) example would be the experiments of Merkel and Hick, discussed earlier. In the assemblage model, the two set-ups are connected as a source and a successor: Merkel's experiment is the starting point for Hick's investigation. Yet these two quite similar set-ups adhere to very different conceptual frameworks. Merkel conducted an experiment so as to calculate the "will-time" (*Wahlenzeit*, time of volition,) and to separate it from the reaction time. Hick replicates the experiment, albeit with the aim of measuring "the rate of gain of information" – the time needed for processing a given quantity of information. While Merkel's experiment is based on Wundt's notion of volition and apperception, Hick examines the principles of information processing. Eventually, Hick showed that rate of gain of information is constant and is a (logarithmic) function of the number of choices between which the subject has to make a decision. The principle became known as Hick's law and is considered to be one of the central principles of user experience design.

3.5.6. Modelling bodily relations with technology

The reaction key was meant to provide access to the mind, but being a material technology, it interacted directly with the subject's *body* (not the mind). It literally attached the subject's body to an electrical circuit and was activated by certain body movements. The reaction key was pressed or released, held between the lips, bitten, tied to the ankle, it was touched and spoken into. There was even a smell-stimulator[217] to measure olfactory reaction, which was to be inserted into the nostrils.

At least at the level of words, I have tried to capture in the ontology these various forms of bodily contact with the reaction key. For each assemblage, I have specified the form of the

[216] This can also be traced in the exhibition: [https://key2mind.omeka.net/concepts#volition]

[217] For a description of the smell-stimulator (after Moldenhauer), see Titchener 1905, 161-162. Titchener categorizes this device as part of the reaction keys family.

subject's interaction with the device: namely, which actions of the subject brought the reaction key into action. The resulting list based on the eleven set-ups includes the following actions: *releasing*, *pressing*, *pushing*[218], *tapping*, *finger moving up and down*, *mouth-opening*, *talking*, *saying 'jetzt'*. This list of actions is aimed at bringing the object back into the horizon of its pragmatics, at making sense of the instrument through its use[219]. In the digital archive, the user cannot press the manual reaction key or type on it. Listing the forms of action that the artifact 'prescribed' could offer her at least some idea of the way it performed[220].

By mapping the actions, one can trace the spectrum of action (or use) of the reaction key within psychological experiments: the way the function or, better said, the script of the reaction key varies. In this sense, of particular interest is the simple telegraph (or manual reaction) key, which has repeatedly altered its program of action. In telegraphy, the key acted as a medium for transmitting a message with each press and release being semantically significant. Further, in Hirsch's experiments and the psychological set-ups, its transmissive function was retained, but no longer did it transfer any more than the simplest and single-valued signal of the subject's response. As the reaction keys became more and more common in the psychological experimental set-up, they found all sorts of uses: for example, in Wells' experiment (1908), the manual reaction key neither captured the reaction nor transmitted a thing, instead acting as a typing instrument to measure physical effort.

Not only does this list of actions point to the scripts of the reaction key, it also in some way reflects the (social) notions of corporeality at the time. The reaction key as a technology implies and encapsulates certain ideas about the subject's body: for example, as to what the appropriate forms of contact between the technology and the subject are.

The most obvious example showing the gap between the then and today's conceptions of the body is the holding of a metal key between one's lips. Such an invasive contact of technology with the body is hardly possible in experimentation today: partly, because of the "danger that an electrical shock will punish the subject and destroy his cooperation in the experiment", as Richards and Guilford (1930, 469) note, proposing a new modification of the

---

218 I follow the naming of the action in the text, so, for instance, the distinction between *pushing* and *pressing* is not entirely clear. Once again this draws our attention to the limits of semantic modelling, which deals only with words, or the naming of things, rather than actions or things per se.

219 Here I am referring to Heidegger's well-known distinction between 'presence-at-hand' (*Vorhandenheit)* and 'readiness-to-hand' (*Zuhandenheit*) (Heidegger [1927] 1962, 98).

220 Another way to contextualize an object through its use would be to document the testimonies of its "users", e.g., subject's experience of contact with the reaction keys. The ontology allows the integration of such testimonies (as well as any other forms of documentation), but unfortunately, I have not encountered such evidence.

key with paper instead of metal. In experimental psychology, the reaction keys were inserted into the mouth, eyes, and nostrils. Cognitive psychology apparently practiced less invasive forms of contact: for example, the throat microphone, which was used in the Fitts set-up of the 1960s (Fitts et al. 1963) is simply worn around the neck (rather than inserted into the throat, as one might think by analogy with fin-de-siècle devices).

Since experimental psychology singled out the mind or mental processes as the object of its research, its interpretations of the bodily experience usually remained behind the scenes. Rather than being found in theory (e.g., in Wilhelm Wundt's grand system), these notions of corporeality are scattered over technical descriptions of experiments and specifications of instruments. Since it is the reaction key that directly interacts with the subject's body, its technical descriptions are particularly telling. Consider, for example, the way Edward W. Scripture presents the construction of his re-action key:

> Two hard-rubber slides run on steel guides. The upper slide has a hole to fit the end of the finger. The other has an inclined hole for the thumb, for use when the key is held by the thumb and finger alone. When the key is rested on anything or is held by the other hand, the thumb may be placed against the projecting arm; this arrangement gives a somewhat easier action, as the finger moves more naturally in a plane inclined to that passing through thumb and finger (Scripture, Moore 1893, 88).

In Scripture's account, the design of the reaction key and the corporeality of the subject – a detailed description of finger movement that extends over nearly a page further on – appear to be intertwined and deeply integrated into one another. The key is designed so as *to comply with* the (most "natural" and "easy") movement of the fingers, yet it also *prescribes* certain movements to the subject. The very construction of the key thus embeds a certain idea of bodily movement and the way it is to be performed.

Mapping the list of actions that activate the key is obviously insufficient to fully articulate this corporeal script of technology, but it is at least a way of drawing attention to the bodily-material dimension. What it does reveal, however, is what movements of the subject's body were supposed to open the door to the subject's mind: finger motion, mouth-opening, etc. Capturing those movements, the reaction key appears as a mediator between body and mind as well as between human and technology.

3.6. Mediations

3.6.1. Mediation: between epistemic and technological objects

In the previous two chapters I modeled the history of things, avoiding the intellectual history – that of ideas, theories, and knowledge. In this model, however, things will be related to knowledge, or more precisely to those scientific (Daston 2000) or epistemic (Rheinberger 1997) objects in the study of which they were involved. The rationale for this move is quite straightforward: a residue of science can hardly be conceived without the context of knowledge it helped to produce. The fact that, say, the reaction key was used to study mental processes, attention or visual perception, rather than to convey messages, certainly well worth reflecting in the archives.

Hans-Jörg Rheinberger's distinction between epistemic things and technical objects provided a vocabulary to model this form of contextualization, which immediately came to be substantially more refined and sophisticated than initially planned. In addition, as part of this model, I adopted the concept of mediation as construed in Don Ihde's techno-hermeneutics.

According to Rheinberger, the situation of research (or "experimental system") is constituted through the interaction of epistemic things ("the targets of research, those things about which we would like to know more" (Rheinberger 2016, 270) and an ensemble of technical objects (including scientific instrumentation, technologies, but also concepts, theories, skills). The technical objects are determined and stable within a given system, while the epistemic things are indeterminate, uncertain, put into question. According to Karin Knorr Cetina, the latter are "partial objects" characterized by "a lack in completeness of being", and "the capacity to unfold indefinitely" (Knorr Cetina 2001, 181).

The distinction between the technical and the epistemic, according to Rheinberger, is functional: it is determined only by the function of the object within the experimental system; i.e., the same object can under different circumstances be epistemic or technical. When the knowledge about the epistemic object[221] is stabilized, it turns into technical objects (and vice versa, technical objects can be called into question and be turned into epistemic ones.

It is only through technical objects that it becomes possible to approach the epistemic object: as Rheinberger puts it, "the technical conditions determine the realm of possible representations of an epistemic thing (Rheinberger 1997, 29). I have taken the name for this relation between the two types of objects from material hermeneutics of Don Ihde and Paul

[221] Karin Knorr Cetina renames "epistemic things" into "epistemic objects" to emphasize their nature as opposed to the stability and solidity of *things* we encounter in everyday life (Reinberger himself subsequently uses this wording). Therefore, here and hereafter I use these terms interchangeably.

Veerbek: *mediation.* Ihde, followed by Veerbek, suggests that instruments actively engage in the formation of the reality under study (or "scientific knowledge"), i.e., they *mediate* it. As Ihde notes, "Instruments form the conditions for and are the mediators of much, if not all, current scientific knowledge. They are the concrete and material operators within scientific praxis." (Ihde 1991, 45). In Ihde's terms, technologies in use have a hermeneutic power (Ihde 1990), that is, through technology we "read" and interpret the world: we imagine and describe molecules as they appear to us under a microscope, or envision embryos from an ultrasound image[222].

Yet this mediation, the essential power of technological objects, is negated on display in the archive, when the artifact is only meant to be an object of gaze, but not a *medium* through which one gazes into another object. In order to reflect this dimension of mediation, I propose to include into the archive the inscriptions and visualizations (Ihde 1999) produced by the technological objects in the course of the experimental situation.

3.6.2. Operationalizing mediations

The *Mediations* model emphasizes the links between scientific artifacts and objects of inquiry tracing their mutual influence and their alterations. It is aimed at addressing the following questions: Which epistemic things have been investigated by this or that technical object (e.g., the reaction key)? And the reverse: by means of which technical objects has this or that epistemic thing been studied? How is a given epistemic object shaped (or mediated) by a particular technology? Last but not least, how do epistemic objects, their mediations and the technical artifacts change over time?

The model is built on the opposition of epistemic and technical objects. The epistemic objects form a special class in the ontology (modeled as a subclass of non-human agents). Each epistemic object *is mediated through* some *inscription* (a class already discussed in the previous chapter). *Inscription*, in turn, *is produced through the use of* some object (it can be a scientific instrument, a conceptual object, an information object).

222 On technology mediation, see also Wise 1988.

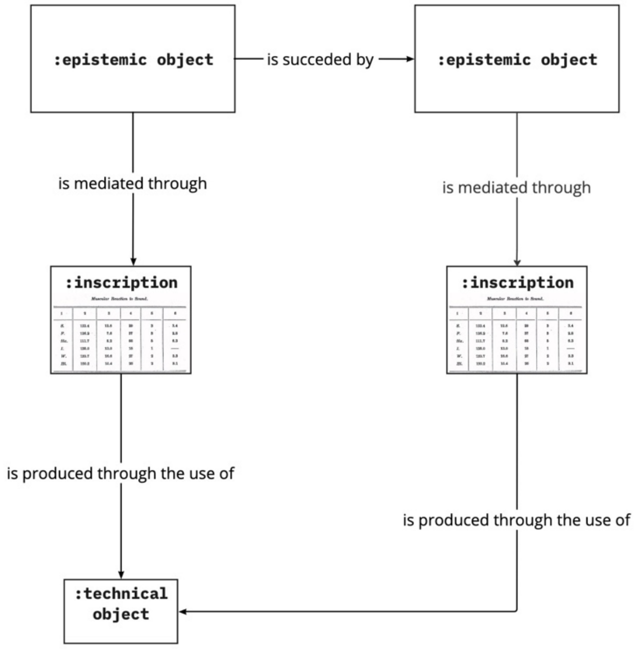


Figure 3.6.1. The general scheme of the "Mediation" model. In this example, one epistemic object succeeds another, but the technical object remains unchanged.

Thus, the model is structurally quite simple: *technical objects* participate in the production of *inscriptions* that mediate *epistemic objects*. The model allows tracking how technical objects change, as well as tracing the transformation of epistemic objects (the changes are modeled using a simple relationship *proceeded by/ succeded by*).

3.6.3. Chasing epistemic objects

As Rheinberger notes, epistemic things are elusive: "In their elusiveness, they stand in contrast to the technical objects – instruments, procedures, apparatus of sorts – with which they are supposed to be brought into interaction. Epistemic objects are thus characteristically underdetermined" (Rheinberger 2016, 270). It is this elusiveness that is perfectly palpable in the attempt to "grasp" epistemic objects and fix them in ontology. The subtle interplay between technological and epistemic objects constantly redefining the contours of the latter is resolutely impossible within ontology and even antithetical to it. To model in ontology means to fix, to define – at least, at some point. The anchoring the epistemic objects within ontology therefore to some extent negates the complexity of Rheinberger's elegant epistemological move.

The semantic model, with its nomenclature of classes, requires a kind of multilevel system of epistemological objects in which more particular issues are united by classes. For example, within the ontology, the epistemic objects "span of attention" and "disturbance of attention" call for being subtypes of the "attention". However, Rheinberger does not outline such hierarchical connections, speaking rather of the dynamics of the epistemic thing: unfolding, vanishment, replacement, and displacement (Rheinberger 2016).

The level of generalization and complexity is defined by the ontology maker in each of the models, but in the case of epistemic objects this choice turned out to be particularly challenging. Although the mediation model registers *facts* (the study of this or that epistemic object by means of this or that technical means), the epistemic objects can take shape in the ontology in different ways, depending on the will of its creator or the task of this or that archive or project.

In this case, my starting point was psychology textbooks and review monographs summarizing the field over time[223]. They allowed me to set forth some general outlines of epistemic objects (as they are considered within the discipline), and following them, to compile lists of relevant publications and experiments that use (and, equally importantly, *do not use*) the reaction keys. I ended up with a range of epistemic objects studied with the reaction keys. The resulting list of epistemic objects appears as follows:

Personal equation,
Reaction time,
Association,
Attention,
Discrimination time,
Reading and word perception,
Mental fatigue.

This inventory, outlining those objects in the study and imagination of which the reaction key was involved, is still remains a matter of researcher's perspective. Each of the objects is potentially fractal and could be broken down into finer objects of inquiry. For instance, in keeping with some (but only some) experiments, I could further divide "reading" into "visual apprehension", "reading speed", "naming", etc. On the other hand, all of the objects listed in turn represent different facets of a single epistemic object – "mind," seen differently at different points from experimental to cognitive psychology.

Further, each epistemic object is traced through a number of publications all the way from experimental psychology to cognitive psychology and is "mediated" through inscriptions

[223] I would especially single out three editions of Woodworth, already mentioned, Boring's History of Experimental Psychology (1929), Whipple's "Manual of mental and physical tests" (1910). For cognitive psychology: Brozek, Pongratz 1980, Gardner 1985.

from the corresponding publications[224]. In the previous chapter, I have already pointed to the importance of inscriptions for contextualizing the scientific residues. In the mediation model, it is through inscriptions that the connection between the technic and the epistemic objects is secured. As opposed to the ephemeral epistemic things, inscriptions could actually be displayed in the archive reflecting the different forms that an object of inquiries take in the course of research situations. A series of successive inscriptions demonstrates how these forms of shaping and conceiving epistemic objects change: say, how psychology moves from mediation of reaction time by means of numbers to a view of reaction time by brain electrical activity, mediated by the event-related potential (Fig. 3.6.2).

TABLE III.[1]

Sound-key.				Lip-key.			
B		C		B		C	
R	R′	R	R′	R	R′	R	R′
164	167	177	176	199	199	172	171
161	159	165	165	185	187	173	173
175	176	175	176	199	201	172	173
170	168	175	172	189	186	177	177
168	168	157	159	166	165	185	176
168	168	170	170	188	188	176	174
19	10	16	10	11	6	13	8

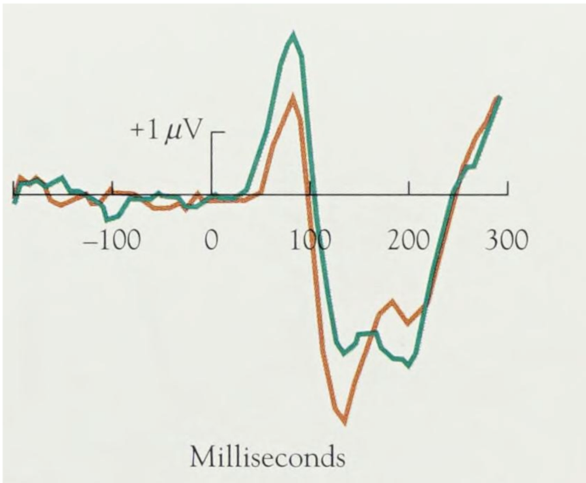


Figure 3.6.2. Inscriptions made by Cattell (1886b, 236) and the visualizations of brain electrical activity (Posner, 1994, 137).

The Mediations model turns the spotlight on two more storylines. First, it makes possible observing the changes of the epistemic objects the reaction key was used to study. Second, vice versa, it allows for tracing the change of the technical arsenal used to study the same epistemic object. In what follows, I will take a closer look at these two configurations.

3.6.4. Modelling epistemic objects change

Let it begin with an example of how the epistemic object is changing. In this case I will take as an example a relatively simple configuration: one in which the technical objects remain relatively unchanged (the reaction keys are used in all cases), but the epistemic object undergoes a transformation.

224 See the corresponding section of the Key2Mind exhibition [https://key2mind.omeka.net/exhibits/show/mediation/technical].

Multiple choice experiments by Merkel and Hick, we have already discussed a number of times, offer one example of almost identical technical set-up being used to investigate fundamentally different epistemic objects. I added a few more examples of similar experiments, revealing a succession of epistemic objects from “will-time” (Merkel 1885) to “rate of gain of information” (Hick 1952) and, following the further development of cognitive psychology to time of different levels of cognitive processing (Posner, Mitchell 1967).

Another, more complex epistemic object to trace is the reaction time itself. From Hirsch and his early experiments at the Neuchâtel Laboratory to Michael Posner, one of the most influential spokesmen for mental chronometry within the cognitive sciences, the “technical” definition of the reaction time remains essentially unchanged. The reaction time is defined as the time latency between the stimulus presentation and the subject’s response. And yet reaction time as an epistemic object – that which is studied – has undergone profound changes.

Still, astronomy is the place to start. Adolphe Hirsch investigated the problem of *physiological time*, which he understood in almost telegraphic terms (Schmidgen 2003), as the transmission of signals from the body periphery to the brain and back:

> […] le but de ces recherches est de déterminer ce que l’on peut appeler le temps physiologique […] temps qui comprend trois éléments […]: 1° la transmission de la sensation au cerveau; 2° l’action du cerveau, qui consiste à transformer pour ainsi dire la sensation en acte de volonté; 3° la transmission de la volonté dans les nerfs moteurs et l’exécution de mouvement par les muscles (Hirsch 1884, 103-104).

Wilhelm Wundt, adopting Hirsch’s experimental set-up and even the very notion of physiological time, nevertheless reframed it, bringing into the psychological realm. For Wundt and early experimental psychology in general, the time between the presentation of a stimulus and the activation of a key was the time of entry of representations into consciousness[225]. “The time taken up by mental processes”, as Cattell (1886b, 63) referred to it, was subject to decomposition into component parts: for example, sensory and muscular reaction time (Lange 1888).

The comeback of the reaction time experiments in the 1950s took place in a completely different system of coordinates. Reaction time was no longer thought of as a sign of the operation of consciousness, but as the processing time of the human nervous system spent on cognitive operations (Posner 1986, 7). Cognitive psychology thematized information

[225] It is noteworthy that even in early experimental psychology and within Wundt’s circle, interpretations of reaction time differed. Some of the debate can be followed in Schmidgen 2014.

processing rather than apperception and will, and problematized the distinction between serial and parallel processing, rather than between sensory and muscular reactions.

The reaction keys, on the other hand, were only pushed out of the experimental set-up when there comes the technology to observe the response of the brain itself: EEG allowing for measuring event-related potentials and later PET and MRI scans. Yet with the advent of these technologies, the epistemic object has also changed drastically: recording the reaction of the brain, PET and MRI scans no longer question the reaction *time*, but instead raise the issue of *space* – those parts of the brain that are responsible for the response (Fig. 3.6.3).

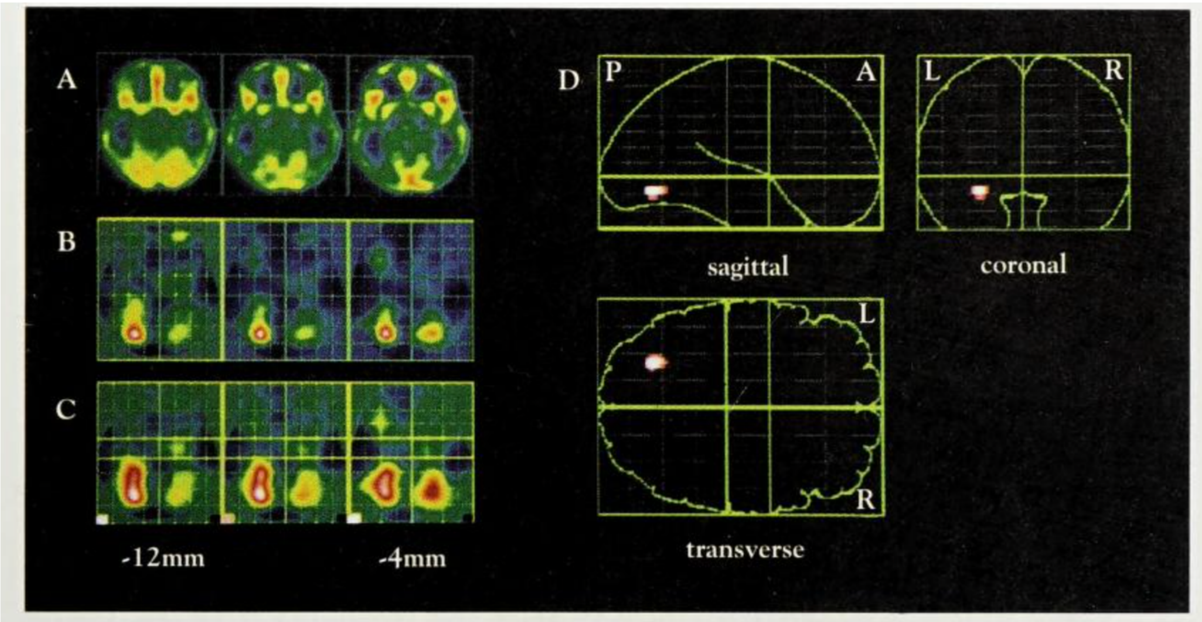


Figure 3.6.3. MRI scans in the reaction time experiments (from Posner, 1994, 75).

3.6.5. Modelling technical objects change

Following Rheinberger's line of thought, in filling out the model I also kept track of those cases where the use of the reaction key was suddenly interrupted, leaving room for some other technology. These cases offer one more ground for making connections within the digital archive: relating those objects that *substitute* each other, providing a sort of paradigmatic context of the artifact. I will offer two examples I have looked at more closely: the studies of association, in which the key and chronoscope were replaced by a simple stop-watch, and the studies of reading, in which the traditional reaction time set-up was replaced by a device for fixing eye movements.

The studies of association offer a good example of how a change in technical means (in this case, the abandonment of chronoscopes and reaction keys) can be indicative of a change in the very ethos of the discipline. Experiments on the measurement of association time began in the Leipzig laboratory in the 1880s and were entirely consistent with the imperative of precision (Benshop, Draaisma 2000) that characterized the approach of the Wundtian school. Its reference point is usually considered to be the now familiar Cattell's experiment with Hipp chronoscope, resulting in a reaction time to the nearest thousandth of a second. By the end of the 1890s, however, the set-up, in which response was captured by the key and measured by the chronoscope, was replaced by a simple stop-watch controlled by the experimenter. In a trade-off between precision and convenience, experimental psychologists have increasingly opted for the latter. Even Cattell himself, a great adept of precision and instrumentality, preferred to use simply stop-watch in large-scale association experiments, arguing that it was inconvenient for a large number of subjects to use a cumbersome and ultra-precise set-up with chronoscope and reaction key (Cattel et al. 1889, 230).

Carl Gustav Jung, in his book on association, put it more conceptually and mocked the paradigm as a whole:

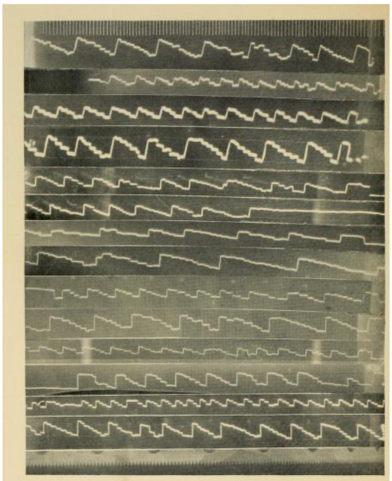
> "So long as we have not sufficient knowledge of the causes of the variations, small differences in the times cannot tell us anything. We do not, therefore, require any complicated experimental conditions to measure times of one-thousandth of a second; we may quietly ignore slight differences so long as the causes of the greater differences do not escape us. Apart from the fact that complicated methods of measuring the finer intervals disclose nothing more than measurements with the one-fifth second stop-watch, there are weighty considerations against the use of mouth-whistles, trumpet-calls, or dark chambers" (Jung 1919 [1906], 228).

This comment by Jung is not only a justification for the use of simpler technical equipment, but also an open disregard for the ideal of precision that Wundt and his disciples so assiduously cultivated. The change of the technical objects unambiguously signals a change of research ideals and reference points. The tendency toward the qualitative study of association rather than its quantitative measurement is finalized by the 1930-40s (Robinson 1932), whereas in the 1950-60s reaction keys and more accurate time-measuring instruments return to the experimental set-ups.

Research on reading and word perception exemplifies a different tendency: that of a technical object being replaced by a technical innovation. The histories of reading research conventionally begin with the very same Cattell's experiments in which he measured the rates of naming of words and letters and concluded that words are perceived in their entirety (Boring 1929, 527-528). Relying heavily on Cattell's results (but also rethinking them), in 1898 Beno Erdmann and Raymond Dodge published one of the first full-scale studies of reading (Erdmann, Dodge 1898) in which they combined two methods: the "classical" measurement of reaction time and the observation of eye movements in reading (which, in the absence of the necessary equipment, they carried out themselves). A few years after their book was published, several eye-movement recording apparatuses appeared (the most popular, photographic, recording technique was elaborated by Dodge himself). Almost immediately eye movement registration completely replaced response tracking by means of reaction keys. Until the very 1940s, all the influential reading studies– the monographs by Dearborn (1906), Huey (1908), Gray (1917) – focused exclusively on eye movement registrations and relyed on them for their reaction time calculations. Naturally, the emergence of the new technique entailed a transformation of the epistemic object: “the reading time” changed its outlines[226] and was now mediated and seen not through numbers, but through photographic records of eye movements (Fig. 3.6.4.).

TABLE XXXVI.

English.				German.			
B		C		B		C	
R	R′	R	R′	R	R′	R	R′
493	484	451	450	419	409	501	498
481	475	490	488	451	454	533	527
447	440	451	457	424	418	507	500
391	383	430	434	379	370	433	432
391	378	431	431	381	376	473	475
441	432	451	452	411	405	489	486
37	21	20	13	31	20	24	15



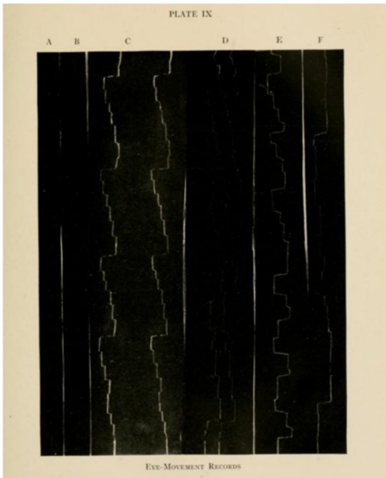


Figure 3.6.4. The Inscriptions made in the reading experiments by Cattell (1886c), Dearborn (1906), and Gray (1917).

3.6.6. Epistemizing the reaction key

According to Rheiberger, in the course of experimental situation, technical objects, stable and determined, at times come into question and turn into epistemic ones. Within the archival context, the episodes of such “transformations” can be quite telling. They reveal that behind a particular residue lies a certain reflection, that it has been itself an object of imagination and inquiry. I have identified a few such episodes in the history of the reaction keys.

Episode 1. Performance analysis

The first time the reaction key became an object of analysis was within the framework of early experimental psychology. The multitude of invented devices and apparatuses called for their

[226] From this point on, the “reading time” was broken down into the time of eye movements and the duration of fixational pauses.

comparison and evaluation (Hill, Watanabe 1894, Chou 1929). In such experiments, it was not the subject's reaction time that was measured, but rather the performance of different types of reaction instruments. Interestingly, what was investigated in this case was not the key technology itself, but rather its interaction with the subject's body. That is, the performance of an instrument was not defined as such, but through its relation to and contact with the subject. It was this dynamic interplay that shaped the very notion of "response".

Episode 2. Symbolical analysis

Another example comes from a very different realm from the field of psychology. In 1937, Claude Shannon completed his master's thesis proposing a "symbolic analysis" of the so-called "electrical switches" [227]. These switches are devices designed to have only one function: opening or closing an electrical circuit. Shannon himself, of course, appeals neither to psychology nor to reaction devices, citing the telephone technologies and industrial equipment of the time as examples. Yet the functional definition he gives to his object of study – opening/closing the circuit, zero or one, – defines the operation of the reaction keys as well.

In his dissertation Shannon showed that these binary operations of the switches could be described in terms of symbolic logics and Boolean algebra, suggesting that the closed electric circle be considered as 0 and the open one as 1. Thus, any circuit can be "represented by a set of equations, the terms of equations representing the various relays and switches of the circuit" (Shannon 1938, 2). This new perspective of analyzing the electric circuits and their switches laid the foundations of cybernetics and computer engineering, and Shannon's text has been called the most famous master's paper of the twentieth century (Gardner 1985, 144).

Episode 3. Human Factor analysis

Finally, for the third time, the reactions keys found themselves under study (alongside various kinds of buttons, toggles, switches and levers) within the Human Factor Analysis, a discipline taking shape after World War II. The origins of the discipline date to the study of pilot error conducted by psychologists Paul Fitts and Richard Jones in 1947[228]. Fitts and Jones found that a large part of the errors was due to the design and layout of aircraft controls, and suggested that designing buttons, knobs, and levers "according to human requirements" would greatly

[227] Shannon's thesis is preserved in the MIT archives and is represented in its digital collection, see [https://dspace.mit.edu/handle/1721.1/11173]

[228] Fitts, Jones, 1947. The original report is preserved in the archives of psychology at the University of Akron: [https://collections.uakron.edu/digital/collection/p15960coll1/id/25524]

reduce the percentage of aircraft crashes. From this point on, it became apparent that diverse controls (the prototype of which were, to be sure, the reaction keys) mediate the human-machine interactions and therefore deserve special research attention. One of the human engineering guides from the early 1960s, for example, includes a thirty-page chapter (Morgan, 1963, 247-280) dealing with the design of buttons, levers, switches, knobs etc.

The roots of the Human Factor Analysis stem directly from the technological reflexivity of early experimental psychology. Experimental psychologists have been pondering on how to eliminate any bias in the technology-subject interaction in order to provide access beyond their limits, to capture the mental processes themselves. The Human Factor Analysis, by contrast, emphasizes the very interaction between device and human, inquiring into how to attain fit between technology and human perception patterns.

Having mediated between machines and minds in the experimental psychology, the reaction key is being redefined within the Human Factor Analysis as a full-fledged actor capable of affecting human actions and choices.

Conclusion

The "experiment" of archival modelling presented in this part may appear confusing and even provocative when viewed against the "traditional" notions of preservation and archiving. The Western tradition adheres to a method of minimum intervention, operating under the assumption that to preserve (e.g., "heritage") is to maintain an object in as unchanged a state as possible. Authenticity here involves both preserving the material essence (e.g., the *original* materials of a building) and maintaining the *original* order (of this materials).

In architecture, there is a notable alternative approach exemplified by the Japanese tradition, which embraces "preservation through continuous repair and remodeling" (Larsen 1994, 19). In this tradition, modifying a building or replacing its materials does not detract from its authenticity. The act of preservation is seen as an evolving process, embracing change and adaptation rather than maintaining the original condition. Permanence, therefore, is not anchored in the materials or the artifacts as such, but is vested in the continuous practices of creation and craftsmanship:

> In order to repair a building, [Japanese] carpenters had to dismantle it completely or partially at regular intervals. When buildings were dismantled, it was convenient during the reassembling of the parts to introduce new members, which were shaped in accordance with contemporary fashion (ibid, 10).

This practice of preservation through dismantling and reassembling anew resonates with the idea of preservation through (re)modelling. In both cases, preservation emerges as a dynamic process of continual reconfiguration and re-contextualization, grounded in the *present* understanding. As such, it also aligns with the pragmatic conception of preservation formulated in the OAIS model: safeguarding not the objects *per se*, but rather the horizons within which these objects are understood. This "dynamic archive" (Ernst 2015, 102), as previously discussed, aligns with the inherent nature of digital data, which become meaningful only through their (re)use. To echo Bruno Bachimont's perspective on this matter:

> Preserving [digital] content means preserving our ability to reinvent and reconstruct it, by equipping ourselves with the critical tools we need to distinguish between the fictitious and the historical, the imagined and the factual[229] (Bachimont 2020, 9).

As a method archival modelling also invites historians to reexamine their relations with both archival objects and theory. By grounding my ontology in theoretical approaches, I have thereby injected theory *inside* the archive. This move, in fact, is not without controversy, as archives are often considered an *antidote* to theory. The original archival object is perceived as

[229] My translation from French.

a reliable slice of reality, whose authority transcends any intellectual construct. As articulated by historian Mary Lindemann:

> theory for historians [...] has more often proved an untrustworthy guide than a reliable pilot. Immersion in the archival evidence seems to me an obvious corrective to theoretical flights of fancy' (Lindemann 1992, 154).

Modelling as a method proposes an inverse configuration: in it, "theory" remains stable (takes the form of the rules of utterance, nomenclature of classes and their relations), whereas the objects (or their representations), conversely, acquire meaning solely in relation to and through the lens of the theory.

Both configurations are not without flaws and harbor a good deal of epistemological pitfalls. The allure of archival records and their "reality effect" has been subject to critique by both historians (La Capra 1985, 92) and archaeologists (Shanks 1992, 99). Both argue this fascination amounts to archival fetishism, where the archival fragment is seen as a substitute for the past itself[230]. Modelling, in its turn, carries the risk of naturalizing or universalizing theory (or model). It overlooks aspects that defy formalization, such as the instability of relationships, the fluidity of categories, the materiality of objects. It is entirely dependent on the creator's will and the "ontological" choices they make.

Modelling thus emerges as a form of "making" the past, and as such, it requires continuous reflection on its constraints, capabilities, and scope. However, if we follow the constructivist tradition in historiography, which posits that the past is never simply given to us as it is but is always constructed, whether through narrative forms and rhetorical figures (White 1986), visual codes (Bann 1986), or the historian's position in time (Danto 1968), then the idea of archive modelling appears as an epistemologically viable (or justified) approach.

As such, it introduces a new (epistemological) configuration, in which the archive appears as a form of historical representation in its own right, alongside the narrative. In other words, the archive does not precede the interpretation of the past but is itself an interpretation.

I see three potential advantages in such a configuration. Firstly, with this approach, the archive turns into the "scene of invention" (Biesecker 2006), a site for generating new understandings and interpretations. Secondly, modelling opens up the possibility of creating *multiple* archives built on different models, thereby multiplying interpretations and histories. Thirdly, modelling can serve as a solution to the issues of object contextualization. Through the process of modelling the archive, as we have observed, it is possible to construct nonlinear,

[230] See also Freshwater 2003.

multifaceted histories, centered around or even driven by *archival objects*, and weaving together things, people, institutions, concepts, and theories. Such a history is open-ended and unfinished, capable of evolving in various directions through the addition of connections. It transcends the confines of conventional narrative structures and can take many forms, such as digital exhibitions to interactive visualizations.

Such a history has the capacity to reveal both the broad strokes of the macro scale and *longue durée* as well as the details of the micro scale. It can be read both closely and distantly, depending on the focus of its creator and the researcher's perspective. It can elucidate, illustrate, complement, correct or even contest existing (linear) historical accounts. Finally, it is able to *operationalize* theoretical concepts and apply them within the archival context (albeit with a number of reservations).

Framing modelling as a method of historical research, this part engages with the epistemology of history and thus is primarily addressed to historians. However, the very idea of archive modelling also speaks to archivists, who have redefined their role over the past decades from custodians to "guides to knowledge"[231] (Cook 1994, 304). Archive modelling, as it seems, could be a method suited for archivists in their capacity as curators of knowledge.

That said, modelling, which simultaneously belongs to the technical, historical, and archival spheres, suggests, or even calls for, varied collaborations between historians, archivists, and knowledge architects. It may also be perfectly suited for ushering in new hybrid roles like historian-archivist-curator, who could finally provide a bridge across the "archival divide" (Blouin, Rosenberg 2011).

[231] Cf. the discussion in chapter 1.1.

Part IV. Making sense of born-digital residues: collective and machine memories of the Stanford Artificial Intelligence Laboratory

> What is in the SAILDART Archive is like what is found in an ancient kitchen midden where broken shards of pottery and tool fragments are embedded in heaps of trash.
>
> Bruce Baumgart[232]

> Some particular draft that was prepared or printed on some particular software, or some particular disk that stores a stage of a work in progress – these are the kinds of things that will be fetishized in the future.
>
> Jacques Derrida[233]

In 1998, John McCarthy, "the father of Artificial Intelligence" and founder of the Stanford AI Lab (SAIL), called on his former colleagues to collect the tapes holding twenty years of the laboratory's backup data. When McCarthy's colleagues, Les Earnest, Bruce Baumgart and Martin Frost, went to retrieve the tapes, they found them in a pile of trash that was about to be taken away. In Bruce Baumgart's testimony,

> We were almost too late, we rescued the tape drive (a full six foot tall rack) from a junk heap at a loading dock of the Paul Allen Building catty corner across the street from the CSD William Gates Building. The junk was about to be recycled (Baumgart 2018, 66).

This "rescue operation" gave rise to the SAILDART archive [saildart.org], exposing the contents of those very tapes – a collection of the Stanford Artificial Intelligence Laboratory digital files from 1972 to 1990.

This anecdote quite vividly illustrates certain mechanisms of transmitting the past into the present. Bypassing all the diligently developed standards for how archival materials should be collected, selected, and preserved[234], archives are sometimes harvested straight from the dustbin of history. The tapes that the former lab members pulled from the junk heap were never designated for archiving: they kept the lab's backup files for the programmers' daily needs, for restoring the data in case of a system crash. SAILDART thus exhibits what by chance did not get discarded, what was not intended as an archive, but at some point, over the years, was realized as one.

---

232 Baumgart 2019, 4.

233 Derrida 1999, 29.

234 The OAIS, as discussed in Part I, provides an exemplary case of such a standard.

Half a century on, SAILDART reflects the life of the institution at the origin of AI and interactive computing. Founded in 1965 by John McCarthy, Stanford Artificial Intelligence Lab, along with the AI labs at MIT and Carnegie Mellon, stood at the cutting edge of a then-new field. SAIL developed and pioneered digital practices and interactive computers, it built robots, automated electronic carts, chatbots, it was the epicenter and heyday of hacker culture. Narrated in this fashion, the history of SAIL as a locus of invention and innovation is well known (McCorduck 1978, Crevier 1993, Nilsson 2009, Nelson 2015). As synopsized by John Markoff,

> Dozens of the world's best computer scientists began their careers at SAIL. More than half of a dozen companies including Foonly, Imagen, Xidax, Vicarm, Valid Logic, Sun Microsystems, Xerox Parc, and Cisco Systems can trace their technology either directly or indirectly to SAIL… SAIL research also led to a wave of AI start-ups in the late seventies and early eighties. (Markoff 2006, 92).

As the Stanford AI lab stood at the origins of interactive computing, SAILDART is accordingly one of the earliest (if not the first) born-digital archives, representing the very beginnings of digital practices. It preserves all that made up Stanford Lab's digital ecosystem: system logs, pieces of software code and its documentation, correspondence and drafts. SAILDART retains both what a responsible archivist or faithful geek would have preserved for the history of computing and what only the lab members would have kept for personal memory, but also what no one would have preserved. While the SAIL history is usually recounted as a history of innovation, SAILDART rather narrates the stories of residual or dead media (Acland 2007; Sterling et al. 2015 [1995]) and electronic waste (Gabrys 2011, Lepawsky 2018).

Situated somewhat between archive and rubbish, between memory and oblivion, SAILDART provides what Mary Douglas terms "a back-door approach" (Douglas 1986, 76) to the history of innovation at SAIL. In practicing this very approach, William Rathje has suggested examining contemporary culture through the garbage it produces (2001, 65), and Jennifer Gabrys (2011) has proposed examining digital media ecology through the history of electronic waste. In a similar move, I suggest looking at the history of the Stanford lab through the lens of its back-up files.

The focus of this part revolves around born-digital residues of science, their (cultural) recycling (Neville, Villeneuve 2002) and their transmission into the present. Historians have extensively discussed the ways of dealing with the new born-digital archives (especially the "Internet archive")[235], literary scholars have addressed writers' electronic archives and the

[235] Brügger 2018, Brügger, Schroeder 2017, Milligan 2019.

peculiarities of electronic text[236], and preservation experts have explored strategies for preserving software[237]. However, SAILDART is not entirely covered by any of these categories. It preserves the digital environment of the scientific institution in its entirety, keeping an extraordinary variety of genres: from more traditional textual ones (correspondence, essays, drafts) to some pieces of software, arbitrary code fragments, undecipherable dump files. Created by AI pioneers at the dawn of interactive computing, all of these types of files require substantial expertise. Yet given that SAILDART keeps everything without distinction, selection or oblivion, it is quite difficult even to review the archive and identify what is worth attaching this expertise to. SAILDART thus brings forward new calls and challenges for historians (of science) and archivists.

Whereas the previous part questioned the power of the digital archive to contextualize historical artifacts, this part shifts the focus to the *born-digital* archive as a medium of knowledge and memory. Instead of establishing connections between archival fragments across time and across collections, in this part, I will focus on a single archive and delve into its depth. My concern, however, will still be with the epistemological potential and boundaries of the digital (this time, *born-digital*) collection. What kind of knowledge is reflected in such an archive? In what way does it differ from other types of knowledge and how can it enrich historical inquiry? How to understand and interpret born-digital leftovers? And how to investigate it as a specifically born-digital form of residues? To these issues, however, will be added another perspective, that of memory. What does the born-digital archive transmit over time? What is its relationship to the collective memory and the community that made it happen?

I will begin analyzing the archive with its archaeology – the SAILDART technical and media history, its hardware and software. The following chapter looks at the mediology of the archive: its mechanisms and technologies transmission. Chapter 4.3. addresses the question of exegesis of the born-digital archive – the ways it can be deciphered, understood, and interpreted. The last two chapters interrogate what the born-digital collection can bring to the traditional forms of commemoration and "human memories". Chapter 4.4 inquires into how a born-digital personal collection can deconstruct the genre of the "famous scientist's biography". The last chapter explores what the digital archive can contribute to the stories of the collective (scientific) identity of the early AI.

[236] Kirschenbaum 2013 and 2016, Hayles 2002.
[237] Acker 2021.

4.1. Archive archeology: hardware, software and technological memory

I will approach the SAILDART with its forensic history or its (*media) archaeology* – through the approach discussed numerous times in the preceding chapters. Following Wolfgang Ernst (2013) as well as Matthew Kirschenbaum (2012) and Bruno Bachimont (2020), I will be primarily interested in the material and media history of the SAILDART, which affects the way we conceive and interpret the archive.

4.1.1. Memory hardware and history of storage

For a computer, to read information is to write it elsewhere.
Wendi Hui Kyong Chun[238]

The tapes found in the junk heap are located midway through the media-technological history of the SAILDART. The custody of these 229 tapes is detailed in a book and archival preview by Bruce Baumgart (2018, 2019), a PhD student at SAIL from 1969 to 1974, and now a guardian and curator of the lab archives.

The 229 tapes are the so-called "*final* tapes," numbered from 3000 to 3228. They were written from the 2984 *original* tapes that had been used for the SAIL computer back-up since 1972. The *original* tapes, in turn, were preceded by short-lived and giant Librascope Discs, which have even been listed in the Guinness Book of World Records for the largest ever hard-disk platters[239]. Costing $330,000 and holding 50 megabytes, these disks began to melt and "forget" even with a slight rise in temperature. The lab ended up suing Librascope, the discs were auctioned off, and some were made into coffee tables[240].

After the Librascope experiment, a much cheaper and more reliable magnetic disks system was installed in the laboratory, which was in use from the early 1970s to the late 1980s. Files were recorded on seven-track tapes using a special back-up software called DART, developed in 1972. In total the DART program recorded about 50 gigabytes of data on the original tapes.

238 Chun 2013, 133.

239 Cf. [https://www.guinnessworldrecords.com/world-records/635942-largest-hard-disk-platter]

240 This story was told by the former laboratory administrator, Lester Earnest, at the 2009 SAIL reunion at Stanford [https://www.youtube.com/watch?v=gg-mIZIAJco]. The comment about disks starting to forget was made by Les Earnest himself, who also owns the coffee table shown in the presentation. The Librascope disk platter can be viewed at the Computer History Museum: [https://www.computerhistory.org/collections/catalog/102682858]

In 1988 Martin Frost, a system programmer at SAIL, wrote a special software to convert the records from the *original* – low-density seven-track – tape to the *final* – high-density nine-track – tape[241]. The conversion was accomplished in 1990. In terms of storage (or what Matthew Kirshenbaum called “forensic materiality” (Kirschenbaum 2012)) this means that the 2984 reels that by Baumgart’s estimation occupied a 20 by 20 feet room[242] had turned into 229 reels that could fit into a few boxes. In terms of the structure of recorded bits (Kirschenbaum’s “formal materiality”), this means that data that was stored as 6-bit characters had now been stored as 8-bit characters.

The next major storage media conversion took place in 1998. At that point Baumgart and his former lab colleagues read the 229 final tapes into 229 “tarballs”, archive files that wrap multiple records and compress them[243]. Each reel has been translated into a separate tar file. That way, the breakdown of the records into 229 separate chunks was retained – but only until 2014, when Baumgart removed the tar wrapping and wrote all the records into a single linear Linux file.

This conversion left quite tangible traces – that of iron oxide dust – on the plates and required considerable time resources, which Baumgart and colleagues allocated from their non-working time (about 60 hours, at the rate of 15 minutes per tape). Seemingly immaterial, memory has its traces, its media, its cost, and is embedded in certain economic relations. From conversion to conversion, these material contours of memory get changed.

Initially the tarballs were written to several SCSI hard drives of 9 GB each. Out of this version Baumgart made many more versions of the archive – backups of the original backup – which he wrote to the various media available to the IBM worker in the late 1990s. Each of these technologies can be the subject of its own media archaeological excavation: CD disks (the entire archive could fit on about 40), DLTtape cartridges, ADSM systems and others. All these systems and media on which SAIL files have been written across the years have different affordances, that is they differ in how data can be written, accessed, retained, and retrieved. Magnetic tape, for example, requires linear search, a sequential scrolling in order to locate the

---

241 On the history of magnetic tape technology for computing cf. Camras 1988, Daniel et al. 1999.

242 “3000 reels of tapes hang on rails, 48 reels per rail, six rails per rack, each rack is six feet tall four feet wide and a foot deep. So ten racks would fit in a 20 by 20 foot room” Baumgart, [https://www.saildart.org/simple/index-book-simple.html], Chapter 9.

243 The tarballs standards are based on a certain economy of storage. Magnetic tape as a storage device writes data in blocks, leaving significant gaps between the blocks, whereas tarball compresses them.

data needed (much like filmstrip), whereas hard drive data can be accessed in any order. The interpretation of the data and its material form are thus mutually influenced by one another[244].

As observed by Matthew Kirschenbaum, the very process of conversion into another medium, *writing to* as opposed to *writing on* (e.g., paper, a surface), “entails a literal as well as a logical displacement” (Kirschenbaum 2012, 87). Each conversion involves *re*interpreting the data in terms of the new technology to which it is being written.

The provenance of the digital born archive appears as a history of such displacements, reinterpretations, and remediations. The material history of the SAILDART is a history of changing technologies, media, and material manifestations of memory. As Baumgart summarizes in his book,

> Over the past 40 years, the 50 Gigabyte quantity of the SAILDART has transitioned from a room sized off-line big-data set of 3000 reels of twelve-inch tape, weighing 2.2 pounds each; into chip size, on which all of the data can now fit on-line inside one CPU main memory address space (Baumgart 2018, 77).

Not only the physical appearance of the archive has drastically changed, but also the cost of its storage media (let us recall that in the late 1960s, 50MB disks costed $300,000) as well as the technology and economy of reading (if in the late 1990s it took many tens of hours to convert tapes, today it hardly takes more than a few minutes to move the archive to another medium).

This entire media and material history of displacements, conversions and transformations, however, is hidden from the eyes of the SAILDART user. As Baumgart (2018, 80) notes, the user of the website is seeing the very same files as a SAIL programmer in the 1970s (Fig. 4.1).

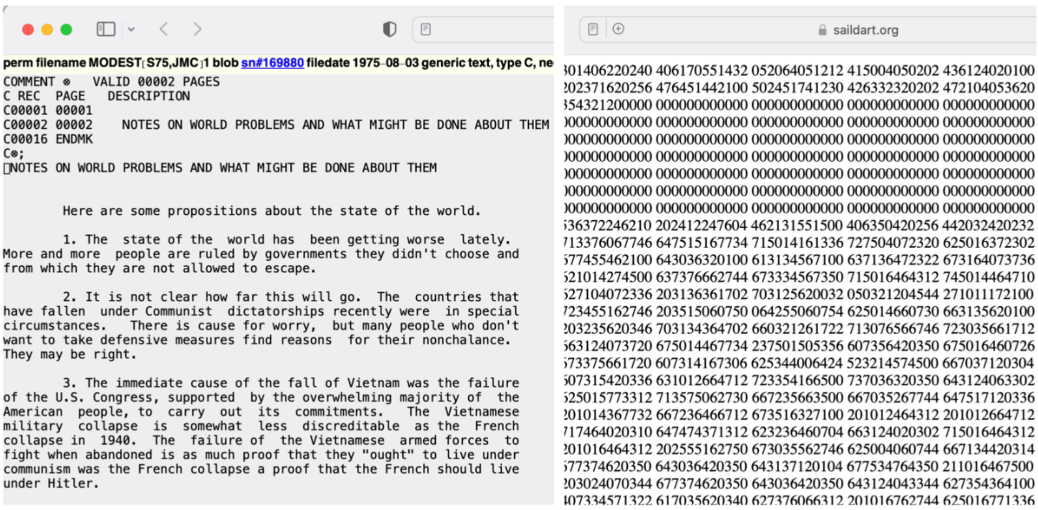


Figure 4.1. An example SAILDART file [https://www.saildart.org/MODEST[S75,JMC]1]. The octal version of the files can also be seen on the website, showing the original bit sequence (right).

[244] For a detailed discussion of this point, see Kirschenbaum 2012, Allen-Robertson 2018.

This effect of direct unmediated contact with the past, *la mise en présence*, that digital technologies make possible has been noted by media theorists. So, according to Bruno Bachimont, the past unfolds through the digital as "a presence that allows us to see and hear not a past that no longer exists, but an almost contemporary reality: the past as if we were actually there" (Bachimont 2021, §2; my translation).

What is important, however, is that this presence effect is made possible due to the numerous layers of hardware conversions and software operations. In order to represent the original bit sequences, the entire medial history outlined above was needed and even more. To this end, the archive had to be translated into encodings and formats suitable for publication on the modern Web (UNICODE, HTML, SQL). Baumgart took it a step further and mirrored the SAIL file system (called WAITS) in a system of web protocols and web addresses. The naming system for files, programmers, and their projects in SAIL-WAITS became the basis for forming the URLs on the SAILDART website[245].

In the case of the born-digital archive – in contrast to the traditional one – the numerous conversions do not challenge the authenticity of the data. Rather, it is precisely the displacements and re-interpretations of "data" that make their very preservation possible. SAILDART's technological history, in this sense, animates the media theories, which recognize the digital files to be agile, non-stable, a "dynamic process between recording and playback" (Bachimont 2020, 46).

4.1.2. Memory software and back-up policies

> These [backup] problems are treated by the technical community as simply the rules of the cosmos.
>
> Theodor Nelson[246]

Since the 1950s computer scientists and especially AI pioneers seeking to make computers intelligent have been concerned about memory. John McCarthy, describing in one of his

[245] The files in the archive, as in the WAITS system, have the following structure:
file name.extension [project, programmer]
For example, in the file https://www.saildart.org/SAIL.MSG[S,LES], SAIL is the file name, MSG is the extension, S is the name of the project, located in the LES (Lester Earnest) programmer's directory. File names in SAIL could be from one to six characters, always uppercase. The extension is optional (three characters after the period). For further discussion of the archive structure, see the next chapter.

[246] Nelson 2009, 63.

interviews his applications for funding to the Advanced Research Projects Agency (ARPA)[247] notes:

> I wanted a large memory… I felt that if one was going to do the things that I had hoped I would soon be ready to do, namely logic-based AI, then I would need a big memory for a big LISP system[248].

Memory management turned out to be one of the conditions for implementing the projects that AI pioneers were dreaming about. This prompted thinking about memory, working on memory exploitation techniques, and constantly searching for memory[249]. Memory was one of the two main assets in the lab (the other one was time[250]). The entire laboratory used one computer, whose memory and time was shared among its members. Each member was allocated "a certain quota of disk space"[251] to store their projects, files and programs.

In the daily practice of the laboratory, memory appear as a matter of preoccupation, accounting, bargaining, and endless negotiation. The archive has a whole directory of memory and disk quota negotiations that have been going on constantly and for many years with the lab administrator, Lester (Les) Earnest[252]. In this directory one finds countless requests for an increase in the quota:

```
∂25-MAR-75 0033        LSP,AJT
 I find that I am very pressed for DSK space. May I have at least another
 100K - I find that keeping 2 or 3 80K FOL core images around soon eats up
 what is available...
 thanks. Arthur[253]
```

```
∂05-MAY-75 0045        CAR,HPM
 I'm managing to live in my disk allocation as far as programs are
concerned,but the aerial photo (which I've chopped up into 16 roughly
standard size pictures each of size 16.1 K) makes this impossible. This is
a request for an additional 250K to hold them, for the duration of the car
finding project.[254]
```

247On the role of ARPA in the development of the early Artificial Intelligence (AI), see Edwards 1997.

248 Interview with John McCarthy. Conducted by William Aspray, Charles Babbage Institute, 2 March 1989. https://conservancy.umn.edu/bitstream/handle/11299/107476/oh156jm.pdf?sequence=1&isAllowed=y

249 For example, in 1959, before founding SAIL, McCarthy developed the Garbage collecting technique, which automatically purged memory that a program no longer needed.

250 On the peculiarities of the time-sharing system see McCarthy1962; Fredkin 1963; Bell, Gold 1972. On the influence of the time-sharing on the structure of SAILDART, see the next chapter.

251 https://www.saildart.org/MONCOM.BH[S,DOC]

252 See, for example, the ACCT.MSG files in the [S, LES] directory.

253 https://www.saildart.org/ACCT.MSG[S,LES]4. Hereafter I quote from the archive documents without alteration of spelling or punctuation.

254 https://www.saildart.org/ACCT.MSG[S,LES]7

```
∂02-FEB-75 0910        ACT,REG
 YOU'D BETTER INCREASE TESLER'S DISK ALLOCATION TO 300K OR FIND ANOTHER
PLACE TO KEEP THE PUB SOURCES.[255]
```

As can be discerned even from this modest sampling of messages, memory in the laboratory was a resource essential to the realization of projects, an object of concern and an object of desire. The use of memory needed to be credibly justified, explained, and substantiated. For example, in the following message Les Earnest informed six programmers at once about the decrease of their disk allocation quotas and invited them to justify their need for more memory:

```
∂25-NOV-75 0108 S,LES          ACH, etc. Disk Allocation
To: ACH, RLL, JAF, JRL, RKN, EJS
CC: REG
I have adjusted your disk allocation to 50k, but will listen to any decent
argument[256].
```

Furthermore, memory allocation became a matter of intellectual policy of the lab. In the face of a permanent memory shortage, the lab's administrators were forced to decide which projects were prioritized by the laboratory and distribute memory accordingly. In the following message, for example, Les Earnest asked McCarthy how much of a priority the development of LISP70, one of the dialects of the List Processing language, was:

```
∂06-JAN-76 1643 LES     LISP70
To: JMC
Dave Smith has requested that his allocation be increased to 300k in order
to carry on with LISP70. This raises a question in my mind about the extent
of our commitment to LISP70. Perhaps you would like to question DAV[257].
```

Since the early 1970s, the laboratory had been struggling with system overload, so memory usage had to be constantly accounted for and monitored by the administration. Exercising control over those who, for one reason or another, overuse computer memory was yet another recurring genre. Following is an excerpt of correspondence between the system administrators:

```
15-OCT-74 1133         ACT,REG
 JMG is eating up the computer, the disk, the UDP and the dart tapes, and
 he's doing it without even being here.[258]
```

And next is an example of Les Earnest's message to a programmer who was caught overusing the disk space:

```
∂25-NOV-75 0059 S,LES          JH
To: JH
It seems to me that for someone who isn't working here, you have an awful
```

[255] https://www.saildart.org/ACCT.MSG[S,LES]2
[256] https://www.saildart.org/ACCT.MSG[S,LES]21
[257] https://www.saildart.org/ACCT.MSG[S,LES]32
[258] https://www.saildart.org/OCTOBE.MSG[ESS,JMC]

```
lot of disk space. Can't you get rid of some of that?[259]
```

In the face of this constant struggle for memory, there was a need for external storage, both to relieve the load on the system, and to ensure the possibility of back-ups in case of a system crash (which was quite common). To this end, in 1972 one of the system programmers of the laboratory, Ralph Gorin, wrote a special piece of software, called DART – Dump and Restore Technique[260]. The program allowed recording files on the (already familiar to us low-density, 7-track) magnetic tape and restoring ('undart') them in the system.

DART was designed for everyday needs. It was based on the pragmatic idea of back-up: restoring the system in case of a crash, rather than establishing an archive; not losing what is needed, rather than saving everything. As Gorin noted in his commentary on the program,

```
The intention of this policy is to provide backup for the disk to
prevent total loss of files in a major crash. A secondary effect is
to provide individuals with short term backup against their own
mistakes. It is not the intention of this policy to provide eternal
backup of every file that ever appeared on the disk[261].
```

Basic DART commands clearly reflected its pragmatics: DUMP (writes files to tape), RESTORE (restores files from tape), and a number of instructions for finding the proper files on tape (LIST, LOCATE, REWIND, ADVANCE, BACKWARD) [262]. This pragmatism of DART is quite consistent with scientific memory practices, which tend to be driven by a single imperative: utility (Georges 2016, 25).

The early versions of DART distinguished three types of back-ups (or "dumps"): full, permanent and temporary:

```
The different classes of dump will treat files differently as follows:
1. RPG and TMP and empty (i.e., 0 word) files will not be dumped
2. Files with protection of 400 or greater will not be dumped.
[…]
3. A full dump will dump all files except those eliminated by 1 and 2
above.
4. A P dump will dump all files (not covered by 1 and 2) that have been p-
dumped fewer than 2 times and are more than n(n=4 at present) days old.
5. A T dump will dump all files (not covered by 1 and 2) that have never
been p-dumped and have never been t-dumped.[263]
```

[259] https://www.saildart.org/ACCT.MSG[S,LES]21

[260] DART had a predecessor – the DAEMON program, whose traces can be found in the archive [saildart.org/DUMP.ADD[UP,DOC]].

[261] https://saildart.org/DART.DOC[TAP,REG]3

[262] https://saildart.org/DART.DOC[TAP,REG]2

[263] Ibidem.

The only files excluded from the back-up were temporary, empty files and those files labeled "private"[264]. All the other files were backed up automatically, depending on how new they were and whether they had been backed up before. All files that were more than four days old became permanent dumps, were saved at least twice and got into the SAILDART. In a way, this policy was driven by computer logic and guided by the system data: the format, the date when the file was created, and the number of its dumps. The backup was done automatically, a programmer could only mark the files that would *not* be recorded on magnetic tape.

This back-up policy, however, has been in flux. The changes to what have been kept and for how long can be traced back to the different versions of the operating notes and manuals in the archive. Thus, the 1972 DART note prescribed the following storage policy:

```
Full dumps will be made at widely separated intervals (probably more than 6
months). Permanent dumps will be made approximatly once a week. Temporary
dumps will be made several times each week.
The retention of the Permanent (and Full) Tapes will exceed 1 year. The
retention of the Temporary tapes will exceed 4 weeks.[265]
```

A decade later, the Monitor Command Manual introduced a special paragraph on the "Dart file Backup policy":

```
The following policy describes the use of Dart for backing up the disk on
tape. There are three classes of system dump: Full, Permanent and
Temporary. Full dumps are never done at SU-AI. (Instead, disk pack copies
are done approximately every 3 months, to protect against disk crashes.)
Permanent dumps are made approximately once a week. Temporary dumps are
done approximately daily Monday through Friday of each week.

Permanent tapes are kept practically forever. Temporary tapes are kept
about one month before being re-used[266].
```

Tracing these variations one can observe how the concepts of short-term and long-term memory, embedded in the DART algorithms, have been gradually taking shape. Since 1982, the horizon of short-term memory has been about a month, the horizon of long-term memory has been almost an eternity. It is notable how a policy that initially explicitly denied "eternal backup" has over time evolved towards eternal storage. One can assume that the very pragmatics of DART has (at least partially) shifted over time: back-up was no longer seen as merely a means of recovering files, but also as a way of preserving some record of what had been created.

---

264 "Protection of 400 or greater" means that the programmer has restricted access to the file for remote logins (login is possible by password), but for local logins the file is available without password. For more information on protection code, see the Monitor Command Manual, Section 5.1. Login. https://www.saildart.org/MONCOM.BH[S,DOC]5

265 https://saildart.org/DART.DOC[TAP,REG]4

266 https://www.saildart.org/MONCOM.BH[S,DOC]26

Similarly, it is quite insightful to follow the changes in the DART commands. To give just one example, the 1978 version of the textbook outlined new DART commands named PUMPKIN and TURKEY:

```
PUMPKIN This command is just like RESTORE except that instead of restoring
files from tape immediately, it queues a request with the Great Pumpkin,
who comes by night to restore your files[267].
```

```
TURKEY This command takes as its argument the number of a DART system dump
tape,and tells you who last used it. It is to find out whose thrug to squuk
if you find a DART tape lying around not where it belongs[268].
```

The very appearance of such commands (as well as their descriptions) shows that the technical *dispositif* of the back-up program is intimately intertwined with the sociotechnical realm of the laboratory. Allowing file recovery to be delayed and implemented at night, PUMPKIN is obviously a response to the system overload, an attempt to decrease the workload of the system during the daytime hours. TURKEY is intended to resolve (probably not uncommon) situations of tape misplacement (as well as to perform an act of social censure – the attitude towards the troublemaker is prescribed by the description). DART algorithms thus appear to be a response to certain social circumstances and settings, embedding certain notions of memory, storage and collectivity.

In practice, however, SAIL programmers were often dissatisfied with DART: its slow pace[269], inconvenience, and the need to scroll through the entire tape to find the desired files. In 1974, Bruce Anderson (DBA) initiated a discussion[270] about the problems and shortcomings of the existing software. As one of the "exercises" to highlight the difficulties, he suggested recovering all computer files for a given day via DART. In his response to Anderson, Mitchell Model pointed out, in particular, the issue of "lack of system self-consciousness", implying that the system failed to keep track of the different versions of a program as it was being worked on. He noted that in this respect "DART helps, but is inadequate".

Despite the twists and turns of the back-up policies, DART did not fulfill the practical needs of programmers. Surprisingly, however, it proved to be perfectly suited for the long-term (or even eternal) storage that Gorin had rejected from the outset. Half a century on, the lab's back-up files, created for purely pragmatic reasons, appear as a full-fledged archive assembled in a uniform and faultless logic of the machine.

---

267 https://www.saildart.org/MONCOM.BH[S,DOC]12

268 Ibidem.

269 www.saildart.org/ACCT.MSG[S,LES]4

270 https://www.saildart.org/SOFT.ANS[SOF,DOC]1

4.1.3. Perfect machine memories and the ideal archive

In concluding the media archaeological overview of SAILDART, I will delineate the *dispositif* of memory that results from SAILDART's hardware and software backgrounds.

The debate on the relationship of memory and technology stretches all the way back to Plato, who famously characterized writing as a *pharmakon* (either poison or cure) for memory[271]. In Plato's terms, writing represents a technology external to memory, which therefore can be either hindered or fostered by it. In the case of digital born archives, however, technology can no longer be disentangled from memory. SAIL laboratory files are produced on a computer, backed up by a computer, and retrieved via a computer. SAILDART thus sets a radically different configuration in which the computer turns out to be the carrier of memory, its mediator and even *the subject of memory work*. I will therefore speak of the *dispositif* of technological or machine memory, referring not to the computer storage capacity[272], but rather to its ability to record, encode, store, and retrieve files.

In the "perfect" machine memory, the lab's back-up files are turned into an "ideal" archive that keeps everything exactly as it was and makes total recall possible. Not only does DART preserve the content of files, but it also generates metadata which appears as nearly perfect provenance. It records the exact time of file creation and modification, its creator, its location in the system, as well as captures other (neighboring) files that are located in the same directory. Is this not a literal realization of the archival principle of "original order"? Furthermore, DART, as an ideal chronicler, even registers the very acts of its own "writing": SAILDART stores, among other things, the memory of the backup events, as well as the DART program itself.

SAILDART, thus, brings us closer than ever to the historian's major dream: the total archive with no room for oblivion, in which everything is preserved just the way it was[273]. Such

---

271 Plato, *Phaedrus*. Cf. Jacques Derrida's famous reflection on it: Derrida 1981.

272 Memory as a property shared by man and machine was first conceptualized by John von Neumann, who in 1946 came up with a computer memory architecture. The very idea in von Neumann architecture was originally adopted from psychology by analogy with human memory (for further details on von Neumann's understanding of memory, see Chun 2008). In the 1960s, by contrast, the notion of human memory was modeled by analogy with computers, being described as long and short-term, internal and external, back-up, and so forth. Cf. (Draaisma 2001). For a more extensive discussion, see Chun 2013 (especially, chapter 4).

273 By invoking the image of the "total", "perfect" or "ideal" archive I do not intend to say that there are no lacunae in SAILDART. Of course, one cannot exclude any gaps in the computer memory, which may be due to the aging of media storage, algorithm failures, or various forms of human intervention. Not to mention the fact that it does not reflect the whole spectrum of non-digital experience. That said, I am pointing to the very tendency towards total archiving as opposed to the logic of archival appraisal,

an archive changes the very outlines of memory: memory is no longer "the present of the past" (Ricoeur 2004, 101), the way the past appears in the present, but rather "a form of delayed presence" (Ernst, 2013, 69). As discussed above, the SAILDART user with no distance or mediation finds on the screen exactly what the lab's programmers used to see on their screens back in the 1970s[274]. Technological memory resists the conventional temporal metaphors: it supports neither the idea of the depth of time, nor the notion of successive archeological layers, with deeper layers depositing ever older residues. Instead, it places everything on the same "synchronic level", without historical distance, just the way it used to be at a certain point in history.

Aside from the temporal structures, memory studies tend to articulate the importance and even *normality* of forgetting in the process of recollection. Oblivion is embedded in the very structure of memory. This is also true for the memory of science: Mary Douglas, for example, refers to the "structural amnesia" (Douglas 1986, 70) of scientific institutions that allows them to *re*invent and *re*discover[275]. Technological memory overturns these insights into the structural role of forgetting: it has no place for oblivion; instead, it presupposes stability, permanence, total recall. Technological memories in this sense contrast with dynamic and "living" memory that is transmitted over generations, that passes from one to another through experience or storytelling[276]. The technical *dispositif* of DART, on the other hand, makes it possible not just to recall the past but to actually bring it back to life, literally to restore everything "back to the way it was". For instance, one can reconstitute, reenact the SAIL computer system as it stood on July 25, 1974 (as Baumgart actually did[277]).

SAILDART implies no historical selection, no distinction between major and minor, important and unimportant, historically valuable and non- valuable. As we have seen, the files to be backed up are selected according to the socio-technical logic of the archiving program. With some exceptions, discussed above, DART backs up all files indiscriminately to magnetic tapes, thereby getting very close to a *total* archive.

---

historical selection or "natural" sedimentation. The gaps SAILDART has are what has *accidentally escaped* total archiving. What has been preserved transcends what might be called traces; it is rather an entire environment from the past.

274 I set aside for now the question of whether today we experience these programs and files in the same way as back in the 1970s. This question will be touched upon in chapter 4.3.

275 On the necessity of forgetting in science, see also Rendall, Weinrich 2004 (conclusion).

276 The mode of memory described by Walter Benjamin in his essay "The Storyteller"

277 https://www.saildart.org/j5/index.html

Consequently, Baumgart, as the curator of SAILDART, has had to intervene in the archive in a number of cases: removing some of the duplicate files, commenting on some directories and 'folders'. But the main thrust of his intervention has been restricting access to programmers' private directory files. This is the point where we encounter *human* memory practices[278]. Many former SAIL programmers are not willing to make their total archive public. Therefore, only the general system files, the files of programmers who have consented to publication, and the records of those programmers who passed away are publicly available in the archive[279].

Taking human memory practices out of the picture, however, we are dealing with the "perfect archive", stored in a perfect machine memory and recorded by a "perfect archivist". The following chapters will address how the past is presented to us in such a machine archive and what such an archive might bring to our "non-ideal" human memories and commemorative practices.

[278] Interestingly, however, these "new" notions of privacy do not prevent nude photographs of some unknown girl taken in a laboratory in the 1970s from being made publicly available in the archive. For a detailed analysis of this case from the point of view of gender theory, see (Bergmann 2019, ch. 1).

[279] Baumgart maintains a kind of census of the former SAIL members: https://www.saildart.org/tontine.html

4.2. Archive mediology

What does the perfect and total archive transmit over time? Addressing this question, this chapter offers some observations and reflections on the SAILDART *mediology*. As defined by Régis Debray, mediology is "the discipline that treats of the higher social functions in their relations with the technical structures of transmission" (Debray 1996, 11).

In this chapter mediology is understood quite loosely and does not adhere entirely to Debray's program. However, I draw from it the very focus on and interest in the problem of (cultural) transmission: the way information is carried over time, the mechanisms of mediation, the technologies and dynamics of collective memory. The mediology perspective will serve to sharpen the question of the particularities of transmission in born-digital archives: what gets retained in the born-digital archive and how is it transmitted to us through such an archive?

4.2.1. Total archive structures

At first glance, the SAILDART structure may seem hierarchical and ordered. Yet, upon close examination, it appears to be an intricate jumble, in which everything is intertwined and even the most basic distinctions are difficult to draw.

The archival structure follows the file system on the laboratory computers (WAITS). WAITS was composed of directories, with each directory belonging to a particular programmer and being named by their username (one to three uppercase letters). Usernames usually echoed programmer's initials: JMC was the lab director John McCarthy, LES was the lab administrator Les Earnest, REG stood for Ralph Gorin, the system administrator known to us for creating the DART program, BGB was Bruce Baumgart, and so forth. In the lab, usernames were not only employed for logging in, but also to denote one's digital (laboratory) identity. The acronyms were commonly used among laboratory members to refer to each other, to sign messages, for mentions, citations, links in internal communication and formed a codified in-house naming system.

Further, each individual directory was divided into projects (in the archive they are called "areas", today one would call them "folders") in which a particular programmer stored their files[280]. In the archive, personal programmer's directories are sorted by the number of the files stored (Fig. 4.2).

[280] The archive also contains the common directories for system files and notes (see chapter 4.5 for further description).

perm filename FACT.TXT[SPL,SYS] blob sn#886186 filedate 1990-

```
MA      Martin Abadi
LIA     Lia Adams
RMA     Rich Alderson
JXA     Jack Alpert
DEA     Doug Appelt
PJA     Paul Asente
JB      Jim Ball
BGB     Bruce Baumgart
BB      Barbara Beeton
RJB     Richard Beigel
GLB     Gianluigi Bellin
CAB     Chuck Bigelow
TOB     Tom Binford
GEO     George Black
PMB     Per Bothner
TRB     Tom Bowerman
RBA     Rosemary Brock
AZB     Andrei Broder
ROD     Rod Brooks
RTC     Ross Casley
CHA     Sang Cha
CC      Chris Chafe
DCH     Daniel Chapiro
FC      Frederick Chow
```

saildart.org

rank	first	final	user	prg	Full_name	files
.	1972-11-01	1990-08-17	SYS_72	SYS	System Files	54675
1.	1974-01-15	1990-08-16	LCS_74	LCS	Leland Smith (d.2013)	30643
.	1972-11-01	1990-08-17	2_1972	[1,2] [2,2] [3,2]	waits special files email message files HELP text files	29130
2.	1972-11-01	1990-08-17	JMC_72	JMC	John McCarthy (d.2011)	14564
.	1972-11-01	1990-08-17	3_1972	[1,3]	Public executible DMP files	13988
3.	1972-11-01	1990-08-16	RWW_72	RWW	Richard Weyhrauch	12670
4.	1976-12-16	1990-08-17	CLT_76	CLT	Carolyn Talcott	11323
5.	1973-10-02	1990-08-17	DEK_73	DEK	Don Knuth	10910
.	1972-11-01	1988-12-02	HE1972	HE	Hand-Eye Files	10439
6.	1974-03-17	1990-08-17	LCW_74	LCW	Curt Widdoes	9472
.	1975-12-09	1990-08-17	ACT_75	ACT	Accounting Data Files	9438
.	1982-04-12	1988-12-02	AVB_77	SMI	Sun Microsystems Inc Files	8867
7.	1972-11-01	1990-08-17	LES_72	LES	Les Earnest	8784
8.	1972-11-01	1988-12-02	JAM_72	JAM	Andy Moorer	8612
9.	1972-11-01	1990-08-17	TVR_72	TVR	Tovar	8269
10.	1975-09-05	1990-08-17	ZM1975	ZM	Zohar Manna (d.2018)	8114
11.	1981-07-18	1990-08-17	RFN_81	RFN	Rosemary F Napier	8094
12.	1972-11-01	1990-08-17	HPM_72	HPM	Hans Moravec	8017
13.	1972-11-01	1988-12-02	REM_72	REM	Robert Maas	7755
14.	1973-10-02	1990-08-16	FWH_73	FWH	Friedrich VonHenke	7663

Figure 4.2. On the left is one of the files decrypting the SAIL login system [FACT.TXT[SPL,SYS]]. On the right is the statistics of programmers by the number of files in their directory [https://saildart.org/top200ranked.html].

According to the statistics, Leland Smith, the head of the computer-music group[281], has the largest number of files in his directory, then goes John McCarthy, followed by Richard Weyhrauch from the Heuristic Programming project[282], Carolyne Talcott, Donald Knuth, Les Earnest and so forth. Hardly any patterns or conclusions can be drawn from such statistics. The programmers worked at SAIL for different numbers of years, some had projects that involved a lot of documentation (e.g., Donald Knuth's TEX project), some kept a lot of news bulletins or bibliographies (e.g., Whitfield Diffie, known for developing digital cryptography).

Yet more importantly, the individual directory of a programmer does not correlate unambiguously with their authorship or "scientific productivity", reflecting only the fact that they stored a certain file in the directory. The very same file could also reside in someone else's directory or be authored by a different programmer. Files were copied, files were shared, swapped, exchanged and most importantly, everyone had (almost) full access to each other's files.

The "promiscuous file sharing" (Baumgart 2018, 41) was rooted in the SAIL technical systems. The laboratory operated a *time-sharing* system, that is, all its members used a single

[281] In 1975, the group formed the independent Stanford Center for Computer Research in Music and Acoustics (CCRMA), but remained affiliated with SAIL.

[282] In the early 1970s, the Stanford AI Lab split into two parts: SAIL and Heuristic Programming Project (HPP) under the direction of Edward A. Feigenbaum. HPP is primarily known for its work with expert knowledge systems, including such projects as DENDRAL and MYCIN. Cf. Buchanan 1980.

computer whose process time was divided among them[283]. The system itself thus presupposed a certain concept of collective computing, as opposed to that of personal computing. Lab members shared a single computer, its memory and time, and although all the directories were individual, each was accessible to all other users of the system. The technical *dispositif* paralleled the ethical imperative: the idea that information should be free and accessible to all, the immutable law of the hacker ethical code, as formulated by Steven Levy (2010 [1984], 28-29).

A system in which virtually any of the programmer's files is available to all other users[284] inevitably results in that files are being created in anticipation of the view of the Other. In the archive, one can glimpse how the boundaries of the private and the public, the individual and the collective were constantly being questioned, established, and re-established. One of the lab's systems programmers, Mark Crispin, for example, included a special pre-notification in his address and phone list file:

```
This file is private and unless you're me, you don't belong here. So if you
value your life as much as I value my privacy, don't fuck with this file[285].
```

This private file, which Crispin seeks to protect from the views of others, thus incorporates the notion of the collective. Even a file for oneself is in dialog with someone else; it preemptively addresses to other users of the system who would read it. Therefore, SAILDART is not a mere summation of personal collections of files and records, as one might think based on its structure. In it, the individual is hardly distinguishable from the collective. Rather than a sum of personal voices, it preserves a polyphony, a multitude, a collective voice of the community.

The archive intermingles not only programmers' voices, but also human-made and machine-made records. Computer-generated files are similarly scattered in individual programmers' directories and are not readily separable from those created by humans. In the archive, machine-made records – from computer-generated music and outputs generated by computer models ("*chatbots*") to automatic reports and memory dump files – stand on an equal footing with the code of the programs that produced them. In this sense, the archive can be said to store both human and machine *voices* (or memories).

To complicate this configuration further, the archive also intertwines labor and leisure, personal and work files – a commonplace configuration now, but rather nontrivial at the time.

---

283 An introduction to the SAIL time-sharing system can be found in the archive itself. See, for example, https://www.saildart.org/MONCOM.BH[S,DOC]

284 Moreover, since the early 1970s, when SAIL joined the ARPANET, the files became available to all other network users through the ftp program.

285 https://www.saildart.org/ADDRES[1,MRC]2

In the early 1970s, SAIL introduced a system of individual terminals per desk, so that each programmer could interact directly and individually with the computer whenever they wanted. Tracing the history of the emergence of personal computing, John Markoff notes that "for a period of several years, SAIL had the only system in the world in which the entire staff had a display terminal on his and her desk, including secretaries" (Markoff, 2005, 90). And again, the technical apparatus of interactive computing paralleled a new sociocultural configuration – the emergence of a hacker culture, in which time spent on computers became a form of pleasure, in which leisure time was computer-related and almost inseparable from work. As a result, the archive conveys to us a heterogeneous landscape – different in each programmer's directory – in which program codes and love cards, administrative documents and computer games, fragments of written prose and scientific reports jumble together (Fig. 4.3).

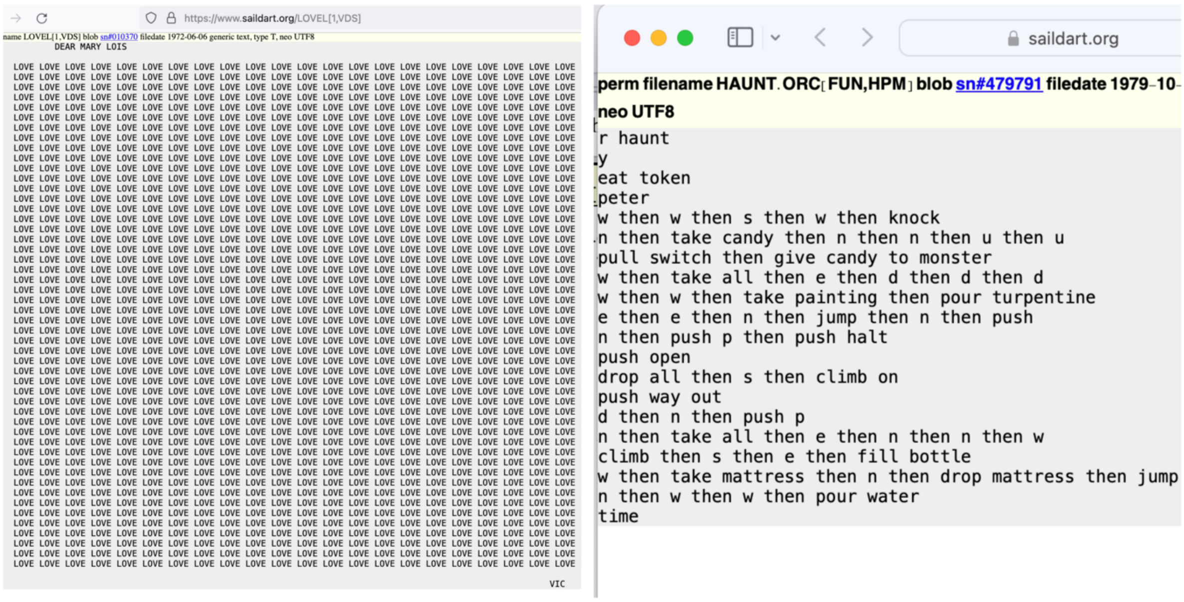


Figure 4.3. Some file examples from the archive: on the left is [www.saildart.org/LOVEL[1,VDS]], on the right is [www.saildart.org/HAUNT[FUN,HPM]].

For an insight into the intricacies of the archive, consider one project from the directory of Mark Crispin, the systems programmer mentioned above. Crispin is primarily known for developing network protocols: he developed TELNET (a teletype network), wrote protocols to connect the SAIL system to ARPANET, advanced a SAIL network called DIALNET, and was the mastermind behind the IMAP (Internet Message Access Protocol).

I will take a brief glimpse at just one, very small project in his directory [1,MRC]. There are only 359 files, the majority of which date from 1977 to 1980. Within this project one finds:

Crispin's code for TELNET[286], news bulletins[287], papers written by someone from MIT[288], fragments of Crispin's correspondence[289] and the beginning of his novel "Software Wars"[290], memos of Network Working Group (not Crispin's) [291], general guidelines for using the lab printer[292], as well as the code for the *Zonker* game[293], a limerick of unknown authorship[294], Crispin's personal statement on nuclear power[295], code snippets from unknown origin starting with someone else's comment: "I HAVE NO IDEA HOW IT WORKS!"[296].

Files by different contributors, the personal and the work-related, limericks and memoranda, games and network protocols, all get mingled even within one very modestly sized project. Note that this is only a small sample of files from only one project by one programmer (Crispin has a total of 26 such projects).

This confusion of genres and the lack of stable attributions of authorship raise questions about the interpretation of the archive today. If any file can end up in any directory and be attributed to any programmer, then how can one possibly draw any statistical conclusions or even treat each directory as a whole, with each of its parts attesting to that whole? Here we encounter the logic of a total archive in which no material has been filtered or sorted out. Such an archive hardly conforms to the principles of the traditional hermeneutic circle. Instead, it exposes the dynamic ensemble of parts and the polyphony of voices.

4.2.2. Future in the past and innovation in reverse

The SAILDART "technological" landscape is no less challenging than its structure. The archive stores files of a great diversity of formats and extensions. Bruce Baumgart provides some statistical insight into this diversity via a word cloud (Fig. 4.4).

286 https://www.saildart.org/TELNET.LST[1,MRC]
287 https://www.saildart.org/AYA.NS[1,MRC]
288 https://www.saildart.org/CRYPT.RLR[1,MRC]
289 https://www.saildart.org/RANDOM.MSG[1,MRC]
290 https://www.saildart.org/SFTWAR.PUB[1,MRC]1
291 https://www.saildart.org/RFC771.TXT[1,MRC]
292 https://www.saildart.org/DDKEY.XGP[1,MRC]2
293 https://www.saildart.org/ZONKER.MID[1,MRC]
294 https://www.saildart.org/LIMER.ICK[1,MRC]
295 https://www.saildart.org/NUKE[1,MRC]
296 https://www.saildart.org/SQRT.MID[1,MRC]

Figure 4.4. Frequency of extensions in the SAILDART (Baumgart, 2019, 11).

The vast majority of these formats and programming languages are outdated, out of use, and unfamiliar even to today's computer scientists[297]. Yet it would be a mistake to assume that in the 1970s and 1980s these languages and formats were in common use. Many of them were originally developed at SAIL and were tailored to SAIL's system: these include, for example, SAIL (SAI extension), a language developed in the laboratory in the early 1970s,[298] or FAIL, an assembler created for PDP-6 and PDP-10 computers[299].

This extraordinary variety of formats stands as a testament to the lack of computer standards in the 1970s and 1980s. Many of these formats originated at SAIL, were in active use for some time, and were than superseded by other formats. So, for example, there are no less than four formats of plain text files in the cloud: PUB, DOC, TXT, TEX.

The unique formats on display at SAILDART constitute but a fraction of the lab's many technological novelties. The SAIL programmers designed for the lab the entire hardware and software infrastructures: from operation system to keyboards, from text editors to fonts, from programmable vending machines (the so-called *Prancing Pony*) to early network protocols (*DialNet*).The laboratory's operating system, WAITS, was one of a kind; the laboratory also had a unique keyboard, featuring mathematical and Greek characters as well as a special modifier key meta, an interactive terminal system, input/output devices, video and audio switches, and a plethora of utilities from text formatting tools to News Service.

297 For a disccussion of obsolete data files and obsolete formats cf. Locklair 2018.

298 Developed by Dan Swinehart and Bob Sproull, SAIL was eventually adapted to the operating system at MIT (called ITS).

299 Ralph Gorin, already familiar to us, participated in the development of FAIL. Other assemblers used in the lab were MACRO-10 and MIDAS.

All these technologies were exceptional and unique in two ways. First, they were customized, crafted, tailored to the unique lab's system, and therefore, as a rule, could not be easily moved to other systems. Second, these infrastructures were designed from nearly the ground zero, in the absence of analogs, in a discipline that at the time had no guidelines, standards, or "good practices". They were developed by the pioneers of a discipline that was about to change the future by making machines "intelligent". As such, all these gadgets and utilities were perceived as radically new media, the technologies of the future. This fascination with the technological future is an indispensable ingredient in the discourse and self-understanding of early AI, which Paul Edwards called "cyborg discourse" (Edwards 1997).

As early as 1980, however, Les Earnest took inventory of the SAIL technologies in the paper entitled "A Look Back at an Office of the Future" (Earnest 1980). Describing SAIL's office work sans secretaries, papers, or clatter of typewriters, Earnest was showing that the visionary project of the paperless office had already become a reality at SAIL.

If as Earnest said in 1980, the future had already arrived, the SAILDART user can witness how what was the future turns into the past. SAILDART houses the remains or fragments of what a half-century ago was "new media". Being archived and brought to the eye of the SAILDART user half a century later, all these "future" cutting edge technologies, however, turn out to be non-functioning programs, non-readable formats and non-operational systems. What we encounter and experience in the SAILDART is not innovation but, as Bruce Sterling (2005, 59) puts it, "innovation in reverse" – obsolescence, outdatedness, disuse. The archive exposes the "pastness" of what not so long ago was the future. Thus, challenging the concepts of novelty, innovation or progress, it makes noticeable the historicity of technologies, gives us the opportunity to experience new media as "historical subjects" (Gitelman 2008)[300].

That said, born-digital archives disclose different modes of obsoletion than do traditional archives. In the previous part, dealing with scientific instruments, we discussed the aging of matter – the romantic patina and mold – which signals the authenticity and value of the artifacts in question. In the case of computer code and software, we are dealing with operational rather than material obsolescence: unreadability of code, non-operability of programs, incompatibility of formats. As Jennifer Gabrys observes, drawing on Benjamin's reflection on history,

[300] Interestingly, many Artificial Iintelligence technologies that were developed back in SAIL – such as chat bots – today still fall under the umbrella of "new media and new technologies" (cf. Acland 2007).

> Instead of demonstrating historical advances, these objects [obsolete technologies] provide evidence of the dust that sediments as a record of these material and technological imaginings. (Gabrys 2011, 104).

Pointing to obsolescence, SAILDART stands in contrast to the histories of SAIL's innovations as often narrated from within the present. In such narratives "innovativeness" is usually seen in the fact that a particular technology used in the lab was a precursor and forerunner of some modern system. Thus, SAIL is often described as a place where proto-blogging, or the first interactive system, or the first computer-programmed vending machine, or the first multi-window interface were invented (e.g. Malloy 2016, Nilsson 2009, Markoff 2006). Such – teleological – narratives are nicely described by Marx's formula "Human anatomy contains a key to the anatomy of the ape", i.e. early forms are made intelligible through more mature forms. In other words, what counts and is worthy of mention in this logic are those technical inventions of SAIL that have led to the "mature" current systems. This is the kind of narrative of progress in reverse, where the past uncovers (only) those inventions that have blossomed in the future.

SAILDART, on the other hand, offers a different – anti-teleological – view of the laboratory's history. It displays all the past "future technologies" indiscriminately: not the inventions deemed innovative from the point of view of today, but at all the systems that were in use at that time – equally obsolete and belonging to the past. Through the born-digital archive, SAIL's technologies and systems are brought to us not as antecedents of today's technologies, but as a bundle of possibilities, a blueprint for a future that may or may not have come true.

4.2.3. PARRY and Prancing Pony: the sedimentation of science in the archives

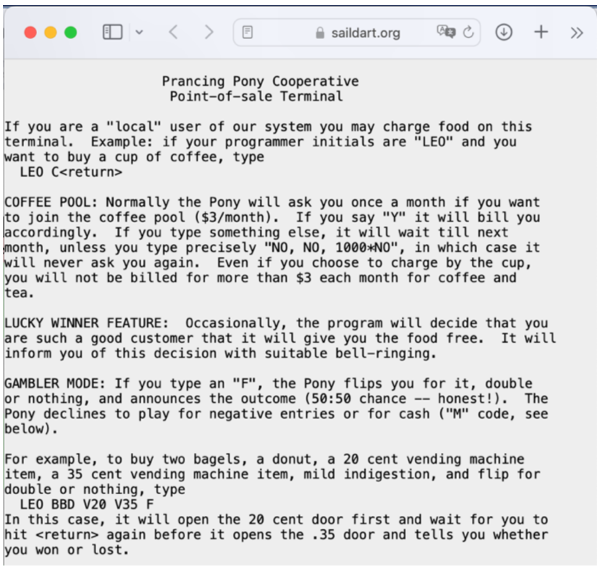
PARRY was the first (and by some accounts, the only) program to pass the Turing test and one of SAIL's most famous programs, known far beyond the narrow AI community of the 1970s[301]. Created by psychiatrist Kenneth Colby, it was a conversational model (or a *chatbot*, as we would say today) mimicking the behavior of a person suffering paranoid schizophrenia. The program was developed based on Colby's theoretical model of schizophrenia, which was operationalized and constructed as an input-output model. PARRY interpreted input messages

[301] See Colby, Weber, Hilf 1971, Mauldin 1994, Saygin, Cicekli, Akman 2000.

written by the user and produced output – context-dependent messages that simulated paranoid communication.

PARRY's closest analog and predecessor was the famous Eliza, created by Joseph Weizenbaum at MIT in the 1960s. Simulating the behavioral responses of a psychotherapist, Eliza was much less powerful. Weizenbaum's model was based on fifty interactional patterns, while for PARRY Colby developed over twenty thousand patterns, which varied according to the levels of "fear", "anger", and "mistrust". Furthermore, unlike Eliza, PARRY was validated by means of a Turing test. Colby had PARRY meet with actual practicing psychiatrists[302] without telling them they were dealing with a program instead of a real patient. To another group of psychiatrists, he gave transcripts of those conversations along with those of the human patients. He alerted this group that one of the patients was the computer program. Neither group was able to determine with certainty which of the answers were generated by the computer. Therefore, it is believed that for the first time, under conditions close to those outlined by Turing in the rules of the "imitation game" (Saygin, Cicekli, Akman 2000), a computer model proved to be capable of "thinking".

Prancing Pony was a "cooperative point-of-sale terminal," or as one would say today, a vending machine. Programmed by Les Earnest, it was the first computer-controlled terminal that enabled users to order food via code and to settle their bills at the end of each month. On top of that, it kept users' credit records, had a gambler mode (issuing double order or nothing), and would occasionally pick a "lucky winner" to be given an order for free (Fig. 4.5).



Prancing Pony Cooperative
Point-of-sale Terminal

If you are a "local" user of our system you may charge food on this terminal. Example: if your programmer initials are "LEO" and you want to buy a cup of coffee, type
LEO C<return>

COFFEE POOL: Normally the Pony will ask you once a month if you want to join the coffee pool ($3/month). If you say "Y" it will bill you accordingly. If you type something else, it will wait till next month, unless you type precisely "NO, NO, 1000*NO", in which case it will never ask you again. Even if you choose to charge by the cup, you will not be billed for more than $3 each month for coffee and tea.

LUCKY WINNER FEATURE: Occasionally, the program will decide that you are such a good customer that it will give you the food free. It will inform you of this decision with suitable bell-ringing.

GAMBLER MODE: If you type an "F", the Pony flips you for it, double or nothing, and announces the outcome (50:50 chance -- honest!). The Pony declines to play for negative entries or for cash ("M" code, see below).

For example, to buy two bagels, a donut, a 20 cent vending machine item, a 35 cent vending machine item, mild indigestion, and flip for double or nothing, type
LEO BBD V20 V35 F
In this case, it will open the 20 cent door first and wait for you to hit <return> again before it opens the .35 door and tells you whether you won or lost.

Figure 4.5. One of the early descriptions of the Prancing Pony [https://www.saildart.org/PONY.OLD[PNY,ACT]1]

[302] In addition to the "sessions" with human psychiatrists, PARRY also "chatted" with a chatbot psychologist, Eliza (Cerf 1973).

As opposed to PARRY gaining a prominent position in the AI history, the *Prancing Pony* occupies more of an anecdotal place in the lab histories. Named after a pub from Tolkien's trilogy, it is usually mentioned somewhere in between the stories about a sauna installed in the lab and the group volleyball games (Levy 2010 [1984], 137, Markoff 2006). It is mentioned as an amusing detail, emphasizing either the playful technological culture of the early AI or Les Earnest's managerial style.

PARRY has a little over a thousand traces left in the archive[303], the majority of which are dialogues with PARRY conducted while debugging the program. Also preserved are a few versions of the source code for the program, files with PARRY patterns, error messages, drafts of Colby's book and articles on Parry, and automated messages sent to users.

Prancing Pony, by comparison, left *thousands* of files spanned over decades. The archive displays invoices, monthly reports, instruction sheets and filling schedules, standard message files that *Pony* should produce for different requests, bills, code snippets, lists of closed and open accounts, lists of debtors, order lists of yogurt and milk (separately for whole, low-fat, non-fat, chocolate milk), reports on bagels, data files, price lists, inventories and stocks, lists of tasks of the "Pony managerial staff", lists of Pony's vendors and other traces of the Prancing Pony accounting system. All these documents are presented in different versions, formats, and for different reporting periods. What is more, the above are only the files stored in a special directory *ACC* (accounting files) [304], and to these can still be added personal accounts and reports scattered across individual programmers' directories.

This example – a deliberately sharpened contrast between the two programs – is intended to capture certain particularities of the digital transmission of scientific residues. Why did Pony leave many times more traces than PARRY did? Why is the vending machine reflected in the archive in greater detail than the first program that passed the Turing test and has been recognized as a breakthrough in the history of artificial intelligence?

This comparison nicely points to the discrepancy between the logics of the scientific history and that of the scientific archive. What science itself recognizes and appraises as significant scientific results is reflected in publications, references, citations, and diverse historical accounts. The archive does not preserve *findings*, because their very extraction from the thick of scientific everyday life requires evaluation, judgement, assessment, making

303 The files concerning PARRY in the Kenneth Colby directory [https://www.saildart.org/KMC] and a special directory https://www.saildart.org/[DIA,SYS] are taken into account.

304 https://www.saildart.org/ACT

distinctions between what is important and what is not. Instead, the archive reflects the surrounding "noise" of scientific everyday life, displaying the residues of material *practices*.

Notably, it is not conceptual results, but material and technical traces of PARRY that remain in the archive: the collection of word statistics and lists of patterns compiled for the model, its source code, iterative dialogs with Parry designed to track whether the model is following the suggested patterns correctly, etc. All the conceptual work, what is recognized as scientific outcome, stays out of the archive, i.e. in publications.

As an everyday and collective practice, *Prancing Pony* turns out to be more important to the living of science at the lab than does PARRY. The vending machine has been embedded in the lab's daily life for decades. Unlike Parry, it was part of a longstanding and collective routine, beyond any particular science project. The same would apply to weekly group volleyball schedules, CPU usage reports, disk quota negotiations, and whatever other routine noise surrounding the production of science.

In a traditional (non-digital-born) archive, PARRY's traces would surely have been preserved in far greater volume than Pony's (at least based on my experience: I have not, prior to SAILDART, come across any (digitized) scientific collection storing cafeteria menus, though they certainly form an important part of the scientific life). The traditional archive involves "historical" selection, sorting what is considered important from unimportant, valuable from non-valuable, filtering out much of this routine material noise of science. In contrast, the digital archive, assembled in the logic of the machine, does not make these distinctions and ends up transmitting a very different picture of scientific life, one in which the everyday and material outweighs what is usually considered to be "scientific". So the "total archive" offers a glimpse into the material and social production of knowledge, into how science is made out of gubbins and scraps.

4.2.4. Computer and collective memories

In all the "mediological" examples above, in one way or another we have come across the theme of the community: SAILDART, as we have seen, transmits or preserves a collective voice (2.1.) and a collective imaginary of the future (2.2.), keeps remnants of collective daily practices rather than those of individual scientific projects (2.3.).

Indeed, SAILDART serves as a form of collective memory. Hardly can one think of a better way to access the SAIL's collective memory, "to gain the group viewpoint, plunge into its milieu and time, and feel in its midst" (Halbwachs 1980 [1925], 118) than to decipher and interpret its digital files. As we have seen, computers stored both the work and leisure, personal and professional lives of the laboratory community. The computer served both as a means of communication within the community and as an object of its imaginary. But more than that, the SAIL community was built around a special relationship with technology and technological expertise. "Computer wizards" of the early AI had a deep understanding of computers: they were intimately familiar with computer architectures, assembled computers by themselves, and were adept at programming in the lowest-level languages. For the early AI practitioners, the computer was not a black box, but, as Sherry Turkle put it, an "evocative object", "an object to think with" (Turkle 2005 [1983] and 2011). In the laboratory where computer served as both a means and an object of research, where the very practices of interaction with and through the computer took shape, the machine memory turns out to be the medium of collective memory, spread among people and computers. So it is the *digital* files that perfectly capture the relationship with and through technology in the SAIL community as it was half a century ago.

In this sense, SAILDART echoes the "Autobiography of SAIL", telling the story of the lab on behalf of the SAIL Time Sharing System[305] (the author of the autobiography was later revealed to be Les Earnest). SAILDART brings Ernest's metaphor to life by offering SAIL's collective history written by a machine. As opposed to any individual testimony, the computer memory is capable of conveying perfectly preserved collective voice exactly as it was half a century ago, without change, alteration, or noise[306].

That said, the external computer memory comes to life and gains power only through the quite traditional practices of commemoration[307] practiced by the SAIL community. These include reunions and celebrations (including giveaways of the Awards for Research Excellence at SAIL in 2009), memoirs, eulogies, and tributes (e.g., Celebration of John McCarthy's accomplishments following his death in 2011), SAIL recollections (Buchanan, Earnest), the community chronicles maintained by Baumgart[308], the SAIL AWAY newsletters that Les Earnest has done over the years and finally, the ever-changing, updating introduction to the archive that Bruce Baumgart is writing. All of these "living memory" practices – commentaries, descriptions, recollections – give meaning to the residues of collective life that the archive preserves.

305 Reprinted in Baumgart 2018, 6-12. The autobiography was e-mailed to former lab members on June 7, 1991.

306 That said, the synchronicity of this collective life can only be conveyed diachronically and discretely by the computer memory *as long as it is read by a human*.

307 On "traditional practices" of commemoration in science, cf. Abir-Am 1998; 1999.

308 https://www.saildart.org/tontine.html

However, the opposite is also true: the archive in turn prompts, catalyzes, and awakens the practices of commemoration. The very maintaining of the SAILDART comes to be a form of nurturing the collective memory: the archive calls for being described, commented upon, curated. In this sense, it is the SAIL community that is the main addressee of the SAILDART. Baumgart, for instance, explicitly addresses his history of the laboratory and the archive primarily to his former laboratory colleagues: he supplies them with “mnemonic exercises” and invites them to fill his memory lacunas with their own recollections. In publications, the archive is mostly referred to by the laboratory members themselves: Nils Nilsson (2009) refers to the archive at a few points in his extensive history of AI, Gareth D. Loy references it in a few articles (e.g. Loy 2013), and of course Baumgart and Les Earnest discuss it a lot.

SAILDART is precious to the community as a testimony to the community itself, but most importantly, as we will see, only the community can properly understand it. It is precisely the question of interpreting SAILDART that I will address in the next chapter.

4.3. Exegesis of the archive

> Future graduate programs will probably have to teach… "digital archeology"…, "digital diplomatics"…, and data mining. In the coming years, "contemporary historians" may need more specialized research and "language" skills than medievalists do.
>
> Roy Rosenzweig[309]

> Unfortunately, most of the software to be read in the SAILDART is painfully primitive and requires a level of expertise similar to that of an Egyptologist who can read 2nd Dynasty Hieroglyphics. I believe there are fewer that 1K persons alive now who can easily read PDP-10 assembly language.
>
> Bruce Baumgart[310]

The SAILDART archive opens the door to tens of thousands of files and probably millions of lines of code. All these files are written in different languages, most of which are already out of use, created for different systems, machines and applications. As with many other born-digital collections, SAILDART thus quite literally and poignantly raises the question of understanding. How to decipher these millions of lines of code written in, as Baumgart justly observes, *ancient* languages? And how can a present-day researcher concerned with the history of the lab decrypt and make use of all these files? In order to show how interpretation can work in the born-digital born archives, in this chapter I will discuss some ways of understanding SAILDART files[311].

4.3.1. Archival recursion, or the key is in the code

> Ein Satz sei mir in einer Chiffer gegeben und auch ihr Schlüssel; dann ist mir natürlich in einer Beziehung alles zum Verständnis des Satzes gegeben.
>
> Ludwig Wittgenstein[312]

A researcher dealing with obsolete SAILDART formats is akin to a paleographist approaching an ancient manuscript without knowing its language. In the absence of a code to the cipher (dictionaries, rules of grammar, alphabet, etc.), they have to proceed from the manuscript itself, looking for the key within the text to decipher. Likewise in the case of the SAILDART, one way to proceed is from within the archive itself, following the ways in which files in the system relate to each other, define or explain one another. In a sense, this approach to interpretation is

---

309 Rosenzweig 2003, 758.

310 Baumgart 2019, 6.

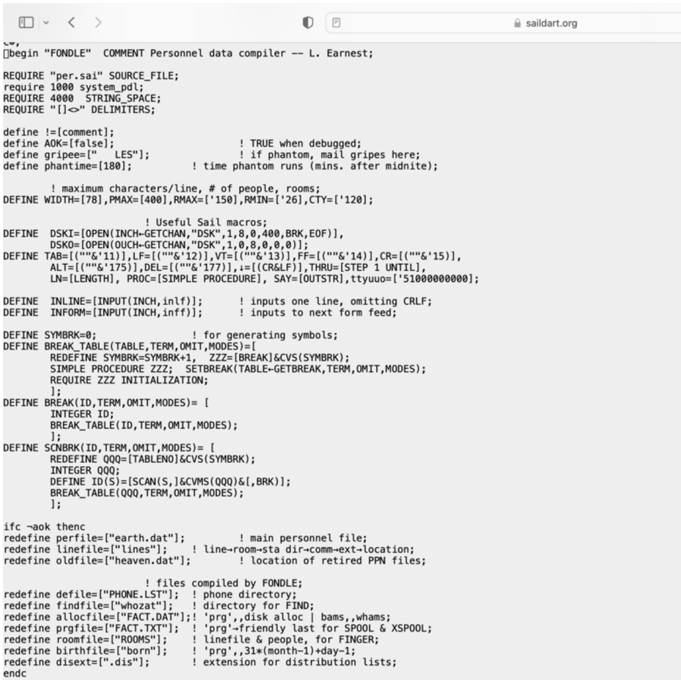
311 On the hermeneutical approach to computer code, see Marino 2020.

312 "A sentence is given me in code together with the key. Then of course in one way everything required for understanding the sentence has been given me". Wittgenstein, Zettel, § 74.

nothing more than the most traditional exegesis as practiced in relation to religious texts. It suggests that all the answers are to be found *within* the Scripture, or in other words, that the key is in the code.

In the case of the SAILDART archive, this means that the key to decrypting any of its files can be found in some other file within the archive. The computing environment is arranged so that some programs produce data for other programs; some files are produced, defined, explained by other files. Hence the archive replicating the computer environment, refers to itself and explains itself.

By way of example, I will use the SAIL system of personnel accounting and its files scattered across different directories of the archive. The starting point will be FONDLE[313] – "Personnel data compiler" – the software to work with personnel files (Fig. 4.6). By the file extension of the program one can identify the language in which it is written – SAIL – and find in the archive a manual on the syntax of this language. The FONDLE utility reads certain files and uses them to generate more files, which in turn are used by other programs. By carefully following this sequence of reading and compilation, one can get a sense of the use and genealogy of multiple files in the archive all at once, as well as take a closer look at the bureaucratic arrangements at SAIL.


```
begin "FONDLE"  COMMENT Personnel data compiler -- L. Earnest;

REQUIRE "per.sai" SOURCE_FILE;
require 1000 system_pdl;
REQUIRE 4000  STRING_SPACE;
REQUIRE "[]<>" DELIMITERS;

define !=[comment];
define AOK=[false];                     ! TRUE when debugged;
define gripee=["   LES"];               ! if phantom, mail gripes here;
define phantime=[180];          ! time phantom runs (mins. after midnite);

        ! maximum characters/line, # of people, rooms;
DEFINE WIDTH=[78],PMAX=[400],RMAX=['150],RMIN=['26],CTY=['120];

                        ! Useful Sail macros;
DEFINE  DSKI=[OPEN(INCH←GETCHAN,"DSK",1,8,0,400,BRK,EOF)],
        DSKO=[OPEN(OUCH←GETCHAN,"DSK",1,0,8,0,0,0)];
DEFINE TAB=[(""&'11)],LF=[(""&'12)],VT=[(""&'13)],FF=[(""&'14)],CR=[(""&'15)],
        ALT=[(""&'175)],DEL=[(""&'177)],↓=[(CR&LF)],THRU=[STEP 1 UNTIL],
        LN=[LENGTH], PROC=[SIMPLE PROCEDURE], SAY=[OUTSTR],ttyuuo=['51000000000];

DEFINE  INLINE=[INPUT(INCH,inlf)];      ! inputs one line, omitting CRLF;
DEFINE  INFORM=[INPUT(INCH,inff)];      ! inputs to next form feed;

DEFINE SYMBRK=0;                ! for generating symbols;
DEFINE BREAK_TABLE(TABLE,TERM,OMIT,MODES)=[
        REDEFINE SYMBRK=SYMBRK+1,  ZZZ=[BREAK]&CVS(SYMBRK);
        SIMPLE PROCEDURE ZZZ;  SETBREAK(TABLE←GETBREAK,TERM,OMIT,MODES);
        REQUIRE ZZZ INITIALIZATION;
        ];
DEFINE BREAK(ID,TERM,OMIT,MODES)= [
        INTEGER ID;
        BREAK_TABLE(ID,TERM,OMIT,MODES);
        ];
DEFINE SCNBRK(ID,TERM,OMIT,MODES)= [
        REDEFINE QQQ=[TABLENO]&CVS(SYMBRK);
        INTEGER QQQ;
        DEFINE ID(S)=[SCAN(S,]&CVMS(QQQ)&[,BRK)];
        BREAK_TABLE(QQQ,TERM,OMIT,MODES);
        ];

ifc ¬aok thenc
redefine perfile=["earth.dat"];         ! main personnel file;
redefine linefile=["lines"];    ! line→room→sta dir→comm→ext→location;
redefine oldfile=["heaven.dat"];        ! location of retired PPN files;

                        ! files compiled by FONDLE;
redefine defile=["PHONE.LST"];  ! phone directory;
redefine findfile=["whozat"];   ! directory for FIND;
redefine allocfile=["FACT.DAT"];! 'prg',,disk alloc | bams,,whams;
redefine prgfile=["FACT.TXT"];  ! 'prg'→friendly last for SPOOL & XSPOOL;
redefine roomfile=["ROOMS"];    ! linefile & people, for FINGER;
redefine birthfile=["born"];    ! 'prg',,31*(month-1)+day-1;
redefine disext=[".dis"];       ! extension for distribution lists;
endc
```

Figure 4.6. FONDLE source code [https://www.saildart.org/FONDLE.SAI[3,LES]]

313 https://www.saildart.org/FONDLE.SAI[3,LES]

The files to be read and processed by the program are specified in a special FILES document[314]. They are all quite easily findable in the archive: PEOPLE (lists of the laboratory staff), LEECH (lists of the lab's guests, usually prominent AI personalities from other universities), CODES (codes of positions and research groups), MAP (room locations), LINES (telephone numbers)[315] (Fig. 4.7).

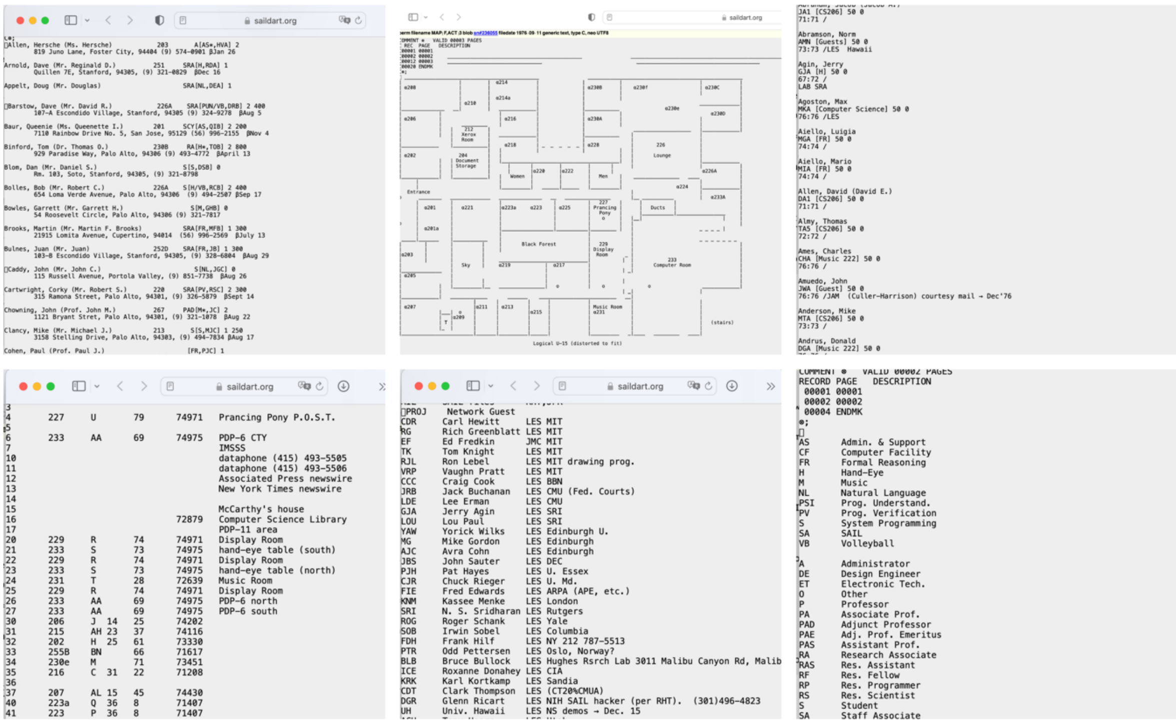


Figure 4.7. Examples of files processed by FONDLE.

All these files remain in flux and are constantly updated, reflecting staff turnovers and changes in their personal data. For example, in 1977, HEAVEN[316] was added to the files read by FONDLE: a list of people who had once worked in the laboratory but had moved on to new positions. At a certain point, SAIL's statistic bureaucracy begins to take into account and keep records of former lab employees, and, as we will see, includes them in its communication networks.

FONDLE reads the above listed files, extracts the data[317], recompiles them, and outputs *new* data records: a phone directory (PHONE.LST), a directory for finding people, a file with

314 https://www.saildart.org/FILES[F,ACT]1. For different versions of the file cf. the directory https://www.saildart.org/ACT.

315 All listed files and their versions for different years can be found in the directory https://www.saildart.org/ACT. The files used in the collage are from 1976.

316 https://www.saildart.org/HEAVEN[F,ACT]2

317 Since 1977, a special program PER.SAI (https://www.saildart.org/PER.SAI[F,ACT]1) is used to read the files and extract data.

disk allocation data (FACT.DAT), a list for SPOOL & XSPOOL (FACT.TXT), a room file (ROOMS), and a file with birthday data (BORN). All these files are in turn scattered throughout different archival directories and exist in many different versions over the years. Tracking their genealogy, one can establish their origins, identify them as machine made, and trace back their sources. Furthermore, some of these files – e.g., FACT.DAT or BORN[318] – store content in octal format and therefore cannot be interpreted at all without the context of the system. It is only by retracing their origins in this manner that one gets to know that BORN stores birthday data and FACT.DAT contains disk allocation data.

Subsequently, these generated files are read by the other programs, which in turn will output further records. Among these "next iteration" programs are: MAIL (sending messages to other users), FINGER[319] (network user program), SPOOL & XSPOOL[320] (scheduling programs for printing files), FIND (locating information or a file), DSKUSE (computer disk usage statistics), BUREAU[321] (computer time usage statistics), RSL (making service reservations). The new files produced by the program are thus used for laboratory members to interact with each other and search for data, they produce file export for printing, and generate different types of statistics (computer time, disk usage).

To avoid going into a bad infinity, I will only take a closer look at only one of the listed programs: FINGER, a utility frequently mentioned in the SAIL histories. Created by Les Earnest, FINGER is considered to be one of the first social networks (Malloy 2016, 9). Exceptionally popular at SAIL, it further expanded across the entire ARPANET network[322], and eventually evolved into Finger User Information Protocol (Zimmerman 1991).

Named after the act of pointing, the program showed the current status of a user in the system. It listed the users who were active at the moment and their exact location; for those users who were not currently logged in, the program showed when they had last been active, as well as their "plan file" – a message they had left for those who would be inquiring for them. "Pointing the finger", lab members could make quick contact, arrange a meeting, find out each other's where-abouts. FINGER also served as an identification program: it decoded usernames, determined a user's location, and gave other kinds of user information.

318 https://www.saildart.org/BORN[PER,ACT]1
319 https://www.saildart.org/FINGER.LES[UP,DOC]1
320 https://www.saildart.org/SPOOL.REG[UP,DOC]1
321 https://www.saildart.org/BUREAU.LES[UP,DOC]1
322 https://www.saildart.org/FINGER.SAI[F,ACT]11

Tracing the genealogy of the files that FINGER uses to respond to user requests brings two points to light. First, this "method" allows us to trace the flow and transfiguration of data: how bureaucratic and formal files (like personnel lists), passed through a computer program, evolved into pragmatic *social* data used for exchange and communication. Following the paths of file transformations one can glimpse how the early institutional-social network of the laboratory was organized. Second, it offers an insight into the genealogy of the technology: for example, the way the FINGER protocol, which has been universally used on UNIX systems stemmed from the bureaucratic (digital) practices and file arrangement within the SAIL lab[323].

Wolfgang Ernst once argued that the digital archive is "self-referent", meaning that its system of links and connections becomes more important than the artifacts it stores:

> The primary operations of the archive are no longer the contents of its files but rather their logistical interlinking, just as the Web is not primarily defined by its contents but by its protocols (Ernst 2013, 84).

In addition to self-referentiality, I propose to address the *recursivity* of (at least some) born-digital archives, referring to the fact that the "total archive" describes, defines, and explains itself. As we have seen, the key to the code is in the code: the archive has everything one needs to understand its files. Proper to computer systems and programming languages[324], the principle of recursion passes into the computational archive, and opens the way for quite traditional exegesis, closely pursuing the hermeneutic circle or better to say the *hermeneutic loop* of the archive.

FONDLE in this sense constitutes one of the most straightforward and simplistic examples. Any scientific project – be it computer music, natural language processing, or robotics – would have a much more ramified and complex file structure, consisting of many components and requiring much more in-depth and sophisticated exegesis.

4.3.2. Machine understanding: AI by means of AI

The above-described method of exegesis, which could be called software close reading, is not always applicable. First, SAILDART stores thousands of files and millions of lines of code. Not all of these files can be examined by such a close-reading. Second, tracing the sequences of input/output and systems of references is often not enough. While this may be sufficient to

---

323 In the late 1980s, several worm programs used the Finger protocol to steal data and infect users' computers, as a result of which the protocol was abandoned for security and privacy reasons. Looking now at the genealogy of the program, this outcome seems quite natural: a program developed within the context of one laboratory, and incorporating its notions of privacy and collectivity, is hardly suitable for global networks.

324 For an extended reflection on recursion, see Hayles 2005.

outline laboratory bureaucratic practices, there are many far more complex programs that do not simply generate files. Instead, they program electronic carts, generate music, assemble air pumps, supervise robots manipulating objects, or engage in dialogues with psychiatrists. Third, many SAILDART files require considerable computer expertise. Even identifying the language in which a particular code fragment is written is not a trivial task for someone who is not intimately familiar with the languages of the PDP computers.

One possible way to proceed in this case is to consult members of the SAIL community: i.e., 1970s AI practitioners familiar with the specific system, languages, and practices. In order to understand a few files, the best approach is probably, indeed, to contact the programmer in whose directory they are located, or their colleagues in the same research group. But in order to look around the archive, to see what is out there, what might be of interest, one would have to inquire with different members of the community about each and every file.

Community language often needs to be deciphered, but how to interpret millions of lines of code written in different languages for different systems that are no longer in use? In order to get an overview of the archive, one needs some kind of "automatic understanding" or "automatic translation" from "machine language" to "human language". The computer-based collective memory requires a new computer-based procedure of exegesis.

As a method of decrypting old files from the history of AI, I propose to call the AI itself. To get the old code snippets deciphered – translated into human language – one can make use of modern AI models such as ChatGPT. ChatGPT will serve as an intermediary between the obsolete code and the user of today who seeks to decrypt it.

As an example of how ChatGPT makes sense of the old digital files, I will look at a rather succinct piece of code called SHAPER in the directory of the programmer Marc LeBrun[325]. Although there are some readable comments in this code, the very content of the program and the way it works remains quite obscure. I provided the link to the code to ChatGPT and asked the latter to give a detailed description of what this code is, how this program works, what its input and output are (Fig. 4.8).

[325] https://www.saildart.org/SHAPER.FAI[SYN,MLB]

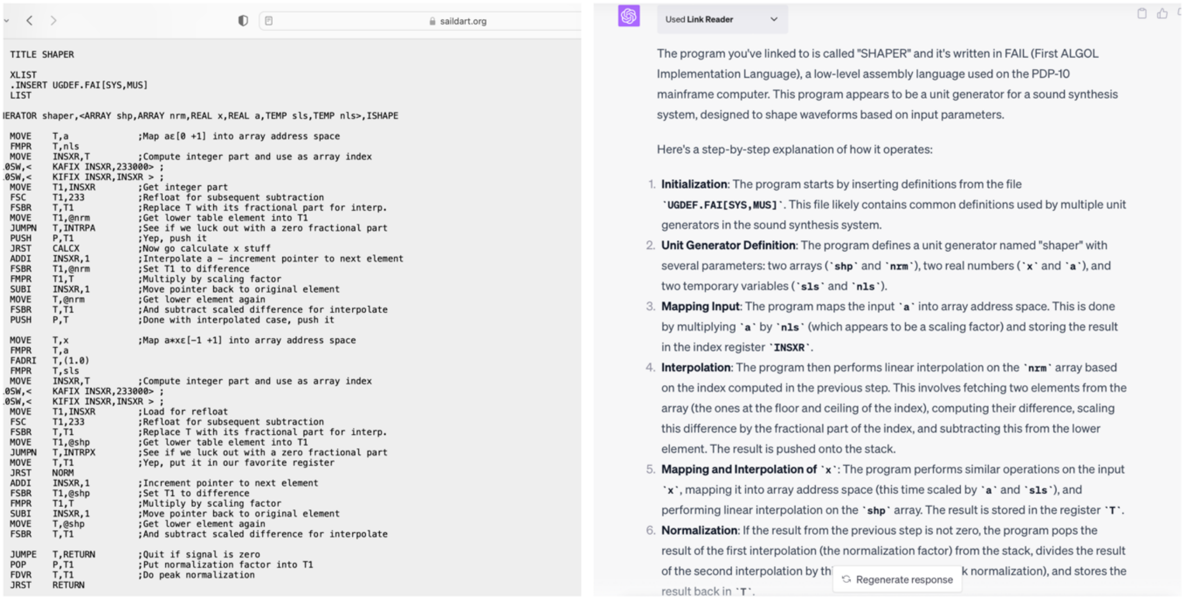


Figure 4.8. On the left is the SHAPER file, on the right is the ChatGPT response.

To start with, ChatGPT is able to identify the programming language in which the code is written (in version 4, according to my experiments it is nearly error-free). Thus, it can be used at least as a first aid in decoding a file (the correct identification of the programming language serves also as a form of verification for further interpretation of the code).

Having defined the language, the model does nothing more than just translating the program algorithm back from the obsolete assembly language to “human” English. ChatGPT sequentially describes all the steps of the algorithm, which can be then compared with the code snippets (the level of detail of the description can be adjusted; one can also ask the model to accompany the description with some code samples). It is thus suitable for exploratory analysis, for getting an idea of what this or that file is about, and how it works.

In the case of SHAPER, ChatGPT identifies the language of the program – FAIL – and its function – to shape soundwaves. The model indicates which other files are referenced by the program and thus makes possible the close-reading as discussed in the previous section. Moreover, it describes the input and output of the program: the fact that numbers are made into music – as well as all those mathematical operations, those forms of distortion to which the sound wave is subjected during the process.

Additionally, in its descriptions, ChatGPT often provides the keywords for further searching, that is, in a way, opens up the code for interpretation. In the case of the SHAPER, it led me to a quest for the “wave-shaping synthesis” – the technology of modifying the shape of a sound wave. Then, proceeding by quite traditional “human” interpretative methods, one can quickly discover that this form of nonlinear sound wave transformation was introduced precisely by LeBrun[326] in the 1979 publication and locate its draft in the SAILDART. Further, one can find out that this method was introduced as an alternative to John Chowning's frequency modulation (FM) synthesis technique, invented at SAIL in the 1960s. Finally, going further, one discovers that Chowning's technique is often cited as one the lab's major inventions, while LeBrun's wave synthesis is mentioned neither in the lab's histories nor even in the studies that specifically focus on the Stanford music research group (Nelson 2015, Loy 2013). While wave synthesis has a certain place in the history of computer music[327], and LeBrun is well recognized as its inventor, the invention itself, however, seems to be quite a forgotten page in the history of the laboratory, preserved only in the computer memory of SAILDART[328].

AI algorithms in this case do a rather literal work of translating from old-computer to human language. The more detailed, specific, close to code is the ChatGPT response, the more useful it is, the more likely it can be verified (It is clear that, in practice, the boundary between "fair use" and blind faith in the "facts" it produces should be clearly delineated).

That said, this very move – using AI as a medium for understanding AI's past – makes sense not just from a purely instrumental or pragmatic point of view. Conceptually, it brings us back to some old debates about the intelligence of machines: in particular, the famous Chinese room argument initiated by John Searle (1980)[329]. This debate goes far beyond the case under discussion, but still deserves at least a brief elaboration.

Searle claimed that artificial intelligence is incapable of understanding (e.g., Chinese). Searle proposed a thought experiment: a person is alone in a locked room and receives notes in Chinese slipped under the door. The person does not know Chinese, but following the instructions of a computer program, is able to produce a pertinent sequence of Chinese characters in response. An outside observer behind the door would think that the subject understands Chinese while the latter was merely following a formal procedure of manipulating characters. Searle's conclusion is that mere manipulation of symbols according to some formal rules is not sufficient for understanding (e.g., meanings or interpretations). As Searle

---

326 LeBrun 1979. Concurrently with LeBrun, the same technique was developed by Daniel Arfib (Arfib 1979).

327 See, e.g., Horner 2007, 64-65.

328 LeBrun has his own "AI avatar" – a chatbot designed to answer questions about his career and expertise [https://www.platohq.com/@marc-lebrun-4e223]. Ironically, this chatbot also knows nothing about wave-shaping technology.

329 For further discussion of the Chinese Room Debate, see Boden 1988, Fodor 1991, Preston, Bishop 2002.

subsequently explains, “system, me, for example, would not acquire an understanding of Chinese just by going through the steps of a computer program that simulated the behavior of a Chinese speaker” (Searle 2010, 17). The implication is that the Turing test is inadequate and cannot prove a computer’s capacity to understand[330].

But what if there are artifacts for which to understand actually mean “going through the steps of a computer program”? Even now (let alone in the future) we are dealing with many objects of knowledge generated by machines and requiring *machine* interpretation of algorithms, signals, bits. SAILDART offers many examples of such artifacts. Getting to understand a program’s code involves precisely simulating its behavior, being able to pursue the algorithm. Classic hermeneutic conceptions of understanding – from human to human – can hardly help us to understand PARRY. In order to understand PARRY, one should apparently be guided not by the model’s responses in a dialog, but by its algorithm – the sequence of steps that it always follows, by syntax rather than semantics. Could it be that in this case it is the artificial intelligence that is the agency capable of *understanding* this particular, already historical, artifact?

And perhaps it is the special AI algorithms that we need to make sense of the ideal chronicles of machine memories?

4.3.4. Putting residues in action: emulations

Moving from human readable to the more and more indecipherable archival records, we finally reach the files that are beyond even the power of modern AI algorithms: memory dumps. These files form a very significant part of the archive: according to Baumgart’s statistics, DMP files are the third most frequent in the SAILDART (after TEX and MSG files). Memory dumps are files that record the state of a system or program at a particular point in time. As a rule, they are used by programmers when debugging programs. Deciphering dumps require expert knowledge of how the working memory of this particular program is configured, moreover, they can only be understood in the context of the entire system in place at the time (again, the unique WAITS).

Dump files present a very interesting case for residue hermeneutics: decoding them requires such a context-dependent and specialized knowledge that after a half-century it appears nearly impossible. In the archival context, the only form of working with DMP files,

[330] Searle’s conclusion sparked a heated debate in the AI community. Some of its echoes can also be found in the SAILDART (cf. the next chapter).

in fact, is putting them in action. If DMPs store the memory state of the system, then the appropriate form of interpretation for them is *re*storing that very memory state at the particular moment in time. Dump files thus provide a perfect example of the performativity[331] of born-digital residues. Like speech acts, they do not describe anything, but can only perform an action, namely execute the memory condition they store. Therefore, the only way of exploring DMPs is emulation – reenacting the original system and software of the Stanford lab.

Software emulation is a well-recognized and elaborated method of preservation, especially in the field of video games studies (Acker 2021, Agnew, Lamb, Tomann 2019). As a preservation strategy, it is usually contrasted with migration of old software to modern formats, media or platforms. Migration involves the conversion of the original software, while an emulator mimics the behavior of an outdated system, making it possible to re-enact and re-animate obsolete software, applications, games in the same way as they operated before. It thus preserves technology through its reactivation and re-enactment. Emulation, in other words, is a way of interpreting computer code computationally[332]. In this sense, emulation seems to be the most natural form of working with old files for the computer community and the most appropriate form of understanding for computer artifacts that were originally meant for operating and performing actions.

It is not surprising, therefore, that Bruce Baumgart addresses the exegesis problem by *recreating* the system, by emulation rather than, for example, *commenting* the code (as a humanities scholar would do). Drawing from the system's DMP file (WAITS) made at 19:04 on July 25, 1974, Baumgart emulated the system as it stood at that exact moment in time (Fig. 4.9). Further within this reconstituted system, he executed some of the software, including SUDS (Stanford Drawing System), GEOMED (design software for 3D models), RAID (debugger for assembly language) and some basic monitor commands. In addition, he also reenacted some of the hardware: the keyboard, the terminal display, the teletype.

---

331 For more on the performativity of code, cf. Mackenzie 2005.

332 This thesis is developed, in particular, by Alexander Galloway, when describing the method of "algorithmic research" (Galloway 2021).

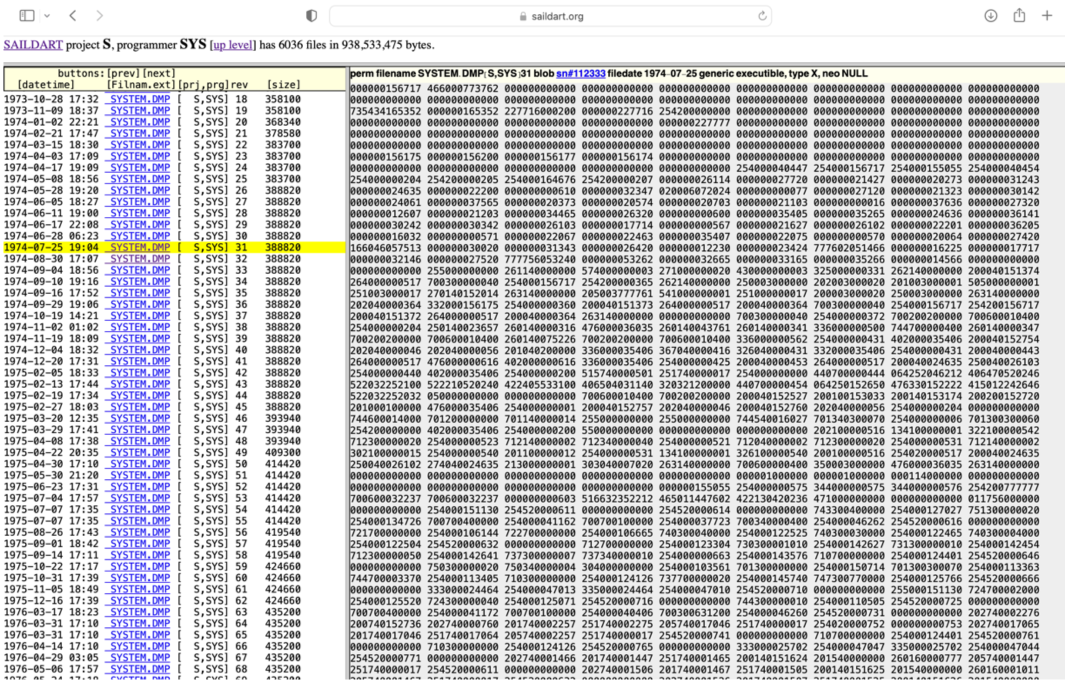


Figure 4.9. DMP file dated July 25, 1974.

Emulation implies encountering the code as action rather than as text. It offers not a hermeneutic understanding but rather tangible experience of interacting with the technology. So, the user can experience the 1974 SAIL system: from the terminal screen displaying the legendary “Take me, I’m yours”[333] and the customized SAIL keyboard (the keys can be clicked with a mouse) to the digital 3D models generated in the GEOMED program.

But even more importantly, as users input commands into the emulated system, they actually get a system’s response. For example, one can input some simple HELLO, DAY or HELP commands and get the system output as it would have been in 1974. Or one can log in as Baumgart [*L 1, BGB*] and inspect the contents of some programmer’s directory as it stood on July 25, 1974 [*DIRECTORY [1,LES]*] (Fig. 4.10).

Emulation takes us back to a specific point in time in the past. Unlike the (textual) archive, it offers no way to trace historical changes (e.g., the ways in which the directory in question was evolving). But what it does offer is the affective experience of contacting the past and interacting with it directly. Emulation gives one a chance to experience and explore the affordances of the technology of the time: for instance, the slowness of the then-new system or its frequent crashing (whether these are effects of emulation or properties of the 1974 system is an open question, though).

333 It is also the name of the “SAIL autobiography” written by Les Earnest (reprinted Baumgart 2018, 6-14).

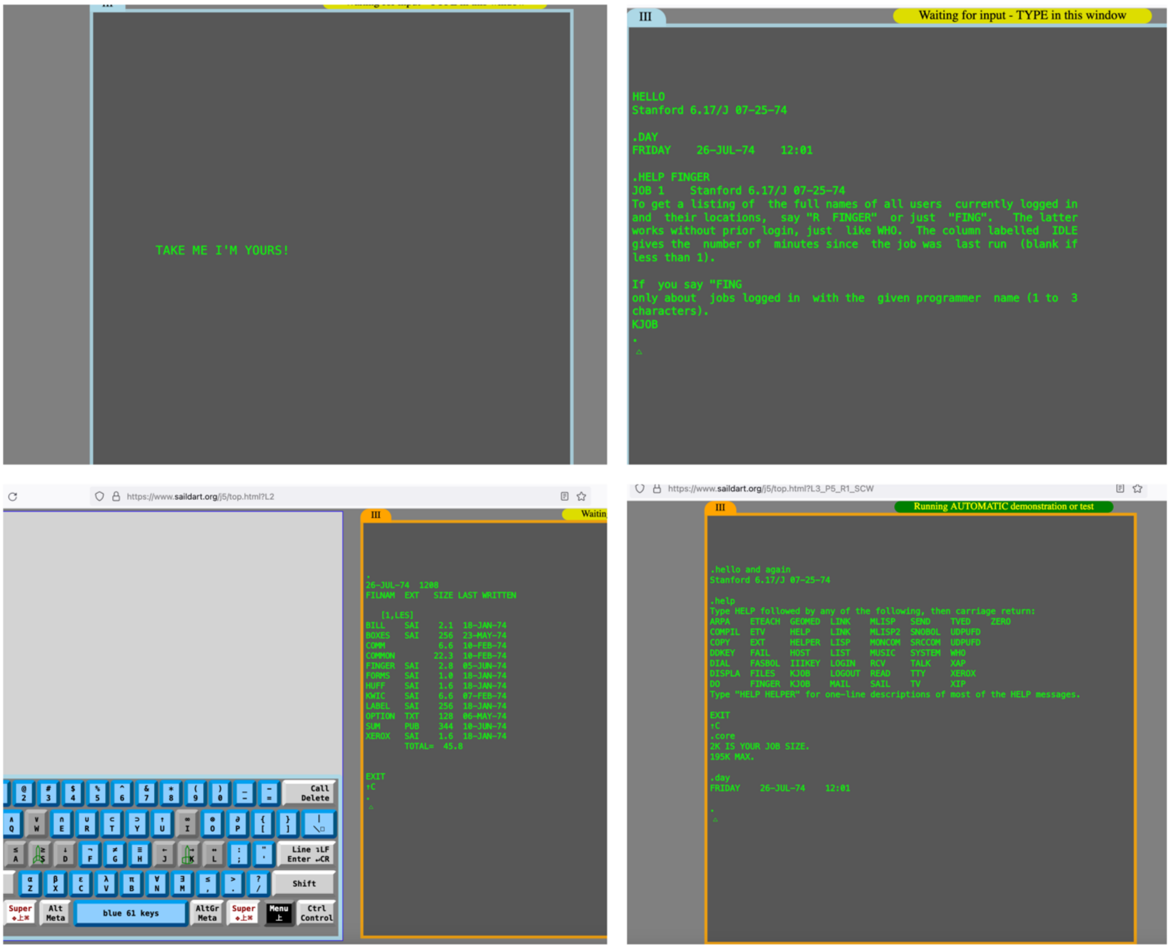


Figure 4.10. Emulation of the system. Top left is the entry view. Bottom left is the contents of the project [1,LES]. Top and bottom right represent executing simple commands .HELLO, .DAY, .HELP.

There are, however, certain limits to this direct engagement with the past offered by emulation. The first and obvious limitation is that the SAILDART visitors encounter an "empty" system not used or shared by anyone. Emulation can only re-animate the technical apparatus, but not *socio-technical* relations and practices. The second and more important limitation for archives of science is that in interacting with the emulated system, the present-day user is still severely constrained by their expertise. Even to experiment with the simplest commands mentioned above, the modern user will likely need to consult the SAIL manuals. Yet, in order to play around with the GEOMED program, the user would be required to possess a far more substantial level of expertise (Fig. 4.11).

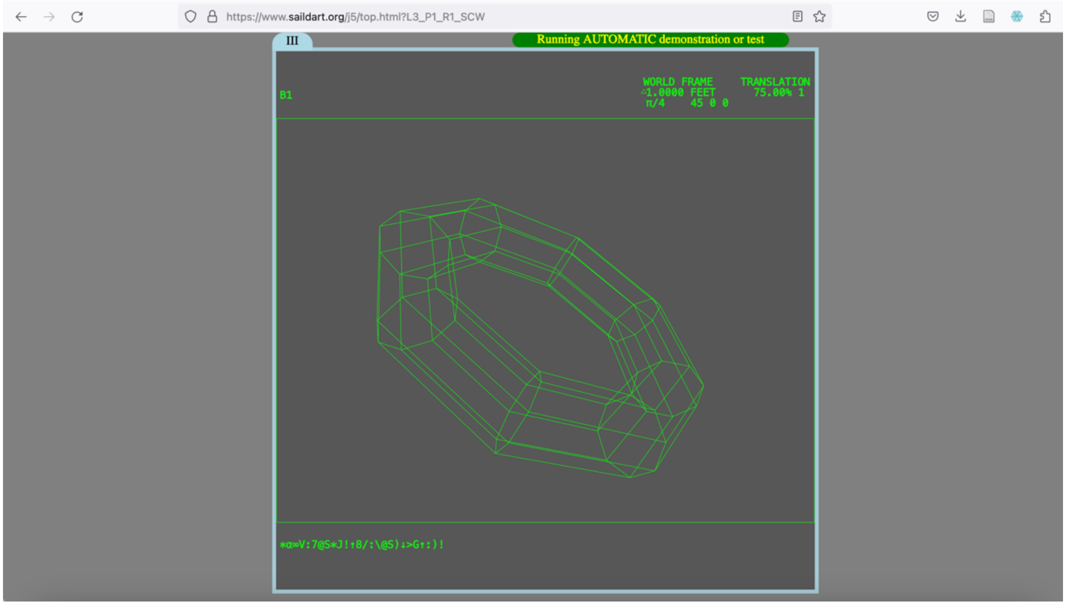


Figure 4.11. A fragment of GEOMED emulation. At the bottom of the image is the code to control the shape.

In the case of video game emulation, the present-day user takes the equivalent/ symmetrical role of a video games player from the past, which usually does not require much skill or knowledge. In the case of emulation of scientific technology, the user is supposed to step into the shoes of an "expert" from the past, but simply simulating the system is not enough for that. Emulation, in other words, retains and emphasizes all the skills and expert knowledge that the SAIL programmers possessed and that the modern user lacks in order to well and truly interact with the system.

Expertise is needed not only to interact with the reenacted system, but also for the emulation itself. In re-engineering WAITS, Bruce Baumgart, according to his testimony, drew on his knowledge of the system and textbooks he had retained from the 1970s. But more importantly, in the absence of documentation, in a number of instances he relied on his own recollections:

> Given that the design drawings for the 1974 generation of the hardware is lost, the only path forward for me is a careful review (a trace) of executing the operating system … and making up fictional representations of the devices I care about based on the interface visible in the code itself augmented with a few paper documents I saved, poloroid photographs I took and memories I can recall (Baumgart 2018, 133).

Reenactment of the system thus requires not only a technical knowledge about it, but also the memory, the community knowledge, the very experience of being in SAIL. Moreover, in the case of SAILDART, the very process of building and running emulation appears as a form of remembrance. In the 1970s, SAIL programmers were developing the "very first and newest" languages and programs. In the 2010s, SAIL programmers are revisiting those very practices and skills to reconstitute a system from the past. Not for nothing, therefore, Baumgart began by reconstituting the very project he himself had worked on as a PhD student in the lab: GEOMED. When emulation is done by the hands of SAIL programmers, the very process of reverse engineering turns into a form of memory, a return to forgotten technical skills, technologies, practices.

4.4. Digital scientific personality: personal collection and scientific biography of John McCarthy

Having successively commented on the material history of SAILDART, its forms of memory/knowledge transmission, and the question of its interpretation, I now turn to some more specific cases. This chapter and the following one compare the born-digital archive more specifically with traditional practices of scientific commemoration, asking what stories does the SAILDART afford and how these stories differ from the existing accounts of SAIL history.

In doing so, these two chapters delve deeper into the history of the laboratory than the previous ones, exploring SAILDART as a source for such history[334]. This chapter discusses the question of scientific personal trajectory by comparing McCarthy's personal directory with his biographies. The next one addresses how the archive sheds light on the SAIL scientific community.

4.4.1. Scientific self: biography and personal collection

In archival and information science, a personal collection[335] is defined as

> a sizable aggregate of an individual's personal traces that the individual or someone else has identified and attempts to manage over time as a relatively coherent unit in order to reflect something important about that individual (Lee 2011, 3).

This rather convoluted definition has three components of interest: first, the idea of "personal traces"; second, the fact that they have been preserved in time through one's efforts; third and most important, that they can reveal "something important" about the person who left them.

The main question of this chapter is what exactly could such traces reflect about the person if 1) the person in question is an (AI) scientist, 2) the traces they left behind are born digital? This chapter focuses on the idea of scientific personality, questioning the ways in which it is portrayed and reflected by the archive of digital residues. What digital residues does the

334 While references to SAILDART occasionally appear in histories of AI (e.g., in Cohen 2017), to date, I have not identified any historical studies that have extensively utilized the archive. As a rule, researchers prefer to use (or cite) other sources if possible. So, for example, in a recent article on the history of robotics at Stanford and MIT in the 1960-70s, Salem Elzway references two documents from SAILDART (Elzway 2023, 159). It is notable, however, that in many other instances he prefers to cite other sources – such as collections of McCarthy's personal papers or his personal websites – even when the same documents are available in SAILDART.

As previously mentioned, references to SAILDART are primarily made by former laboratory members or those closely associated with it, such as Nilsson (2009) and Loy (2013). The archive is also mentioned in passing in histories of computer music, as seen in Zattra (2017), Dahan (2018), and in the history of TEX by Vieth (2007).

335 For a discussion of the term itself and its alternatives ("personal papers", "personal recordkeeping") see Lee 2011, 2-3.

AI scientist leave behind? What might such remnants testify to? And how do they differ from other forms of commemoration of the scientist – in particular, the biography? By way of example, I will draw on John McCarthy's directory, which stands for his personal collection[336], and contrast it with McCarthy's biographies.

In the historiography of science, the problem of scientific personality is thematized primarily in the debates around scientific biography. Tracing the history of scientific biography, Thomas Söderqvist (2007) as well as Michael Shortland and Richard Yeo (2008), show how the individual, the personal, the private are excluded from the historiography of science, which, from a certain point on, favors a sociological approach, the study of collectives, paradigms, and networks rather than individual "accomplishments" [337].

Thomas Hankins (2007), in turn, takes it a step further and argues that the biographical approach is in principle antithetical to the logic of science. As he puts it, "It is ironical that we want to write biographies about just those things that scientists in real life are supposed to avoid in judging their peers" (Hankins 2007, 93). "True science" requires the removal of all that is of interest to biography: personal relationships and passions, religious and political views, social and economic background. Bringing this argument to its logical limit, one could say that an ideal scientist would have no biography at all. Indeed, history is quite aware of such perfect cases: perhaps, Immanuel Kant, whose life was essentially devoid of usual biographical events, came closer than others to such an ideal. The scholar is supposed to have an *intellectual* biography that would reflect "intellectual events" (or events of pure reason) in the biographical subject's life. Hankins takes an elegant move in this sense by suggesting to examine the relationship between biography and the scientific reward system, e.g. patenting. As he shows, the value system of science based on discoveries and inventions "is inextricably woven into the fabric of biography" of a scientist (Hankins 2007, 94). Intellectual biography, in other words, reproduces the scientific reward system.

This "science-based" logic of intellectual biography is readily apparent in the biographical sketches, eulogies, and portraits of John McCarthy. Having read a number of such pieces, one can easily deduce the same "scientific-biographical" pattern followed by the

---

[336] As discussed in Chapter 2, SAILDART personal directories hardly constitute personal collections. However, McCarthy's directory specifically represents an exception to this rule. Since McCarthy cataloged his digital documents, it is safe to say which ones he authored and which ones he did not.

[337] As noted by Thomas Söderqvist, "biography seems to be the ugly duckling in today's discussions about historiographical approaches and science studies methodologies" (Söderqvist 2007, 3). For further discussion of the scientific biography and identity, see also Shapin 1989; Thorpe, Shapin 2000; as well as the Focus section in the journal *Isis* (2006, vol. 97, issue 2).

biographers and interviewers[338]. The underlying leitmotif of all these accounts is, firstly, inventions, and secondly, institutional ties, contacts and collaborations with well-known institutions and personalities[339]. With some slight variations this narrative goes as follows:

John McCarthy studied at Caltech, where, attending a lecture by John van Neumann, he became interested in the idea of thinking computers. During a summer job at Bell Labs, he met Claude Shannon and published together with him a volume on Automata Studies. McCarthy was one of the organizers of the Dartmouth conference, the main "event" in all the AI histories that have been written[340]. Preparing the conference proposal, he coined the very name of the discipline – Artificial Intelligence. Together with Marvin Minsky, McCarthy founded the AI lab at MIT, where he worked on time sharing systems and invented LISP – a list processing language – that for a long time remained the main programming language of AI. It was also at MIT where he came up with "garbage collection technique" and developed the idea of Common-Sense logic – a formalized language to describe common knowledge, which was supposed to be implemented in the Advice-Taker program. In 1962, McCarthy moved to Stanford, where he got a full professorship, received funding from ARPA, and founded SAIL.

Further on at Stanford, there is a remarkable gap in the series of inventions, institutional collaborations, and other conventional events in McCarthy's intellectual biography. About this period it is often said that McCarthy continued to develop the ideas of common sense logic[341]. The only event in McCarthy's Stanford-era "intellectual" biography mentioned in some biographical accounts is the invention of the *circumscription*, a form of formalizing common-sense reasoning. In the logic of biography, almost no event thus falls within the span of time represented by the SAILDART collection. So the first and very pragmatic question that arises is what exactly was McCarthy engaged in during this period? Could the archive potentially contribute some insights to his intellectual biography?

---

338 This biographical account is based on the following biographies and portraits of McCarthy: Hilts 1984a and 1984b, McCorduck 1978, Nilsson 2012, Rajaraman 2014, Shasha, Lazere 1998, Markoff 2016.

339 In describing McCarthy biographers tend to resort to the prototype of the "mad genius professor", socially awkward but brilliant in his work. Most succinctly this image is expressed in Steven Levy's "Hackers", describing McCarthy as: "a distant man with a wild shock of hair and an equally unruly beard [...] A master mathematician, McCarthy was a classically absent-minded professor; stories abounded about his habit of suddenly answering a question hours, sometimes even days after it was first posed to him" (Levy 2010, 11). See also Hilts 1984b, 261 and MkCorduck 1978, 220-222. Among the facts of McCarthy's personal life, usually mentioned are the origin and political preferences of his parents (father – Irish catholic, mother – Lituanian Jew, both activists of the Communist Party), trips to the USSR, the change of political course from the left to the right in the 70's, wives and children.

340 Cf. McCorduck 1978, 93-114, Gardner 1985, 138-155, Edwards 1997, 252-256.

341 See, for example, Nilsson 2012, 8.

I offer a comparison of McCarthy's intellectual biography and his collection not in order to construct a better biography or to contest existing accounts, but to look at the epistemic possibilities of the digital born archives. What are the specificities of the archive as a form of testimony about an individual? What, in comparison with the scientific biography, does the archive testify to? In what relationship to the biography does it stand?

4.4.2. McCarthy's scriptural economy

In 1970, at a conference in Bordeaux, John McCarthy presented the "Home information terminal" project[342]. The paper opened as follows: "Visionaries have often proposed that homes be equipped with information terminals…" (McCarthy 1972, 48). Commenting on it 30 years later, McCarthy noted that by "visionaries" he meant himself and J. C. R. Licklider[343], though his very idea echoed many utopian projects, from avant-garde museums of the future (e.g., Frederick Kiesler's "Telemuseum") to the always-mentioned "Memex" by Vannevar Bush. Like Bush's Memex, "an enlarged intimate supplement to one's memory," McCarthy's home terminal was meant to provide instant access "to files containing all books, magazines, newspapers, catalogs, airline schedules, much additional public information not now kept, and various files personal to the user" (McCarthy 1972, 48).

What is radically different in the two projects is that within a couple of years, at most by 1973, McCarthy actually had such a terminal at his home[344]. Through telephone network it was connected to the SAIL time-sharing computer, and therefore to the ARPANET, SAIL News Service, all its directories and files. Exactly as predicted in his Bordeaux paper, McCarthy just had to type in the name of the file to access it. The JMC directory thus represents the residues of a very early (if not the first) realization of this dream of connecting home to information networks. It reflects the moment when the computer becomes part of the knowledge topography, alongside the laboratory, the bureau, the library, and other topoi of knowledge (Bert, Lamy 2021, 27-146). McCarthy's directory therefore keeps not just

---

342 A draft version of this talk can be found in the SAILDART: https://www.saildart.org/HOTER.ESS[ESS,JMC]3; a version with commentaries thirty years later is published on McCarthy's personal Web site: http://www-formal.stanford.edu/jmc/hoter2/hoter2.html.

343 "I was the main visionary. J.C.R. Licklider was another" [http://www-formal.stanford.edu/jmc/hoter2/hoter2.html], footnote 1.

344 SAIL's 1973 line list of includes McCarthy's house line: https://www.saildart.org/LINE[S,LES]2. According to McCarthy's own testimony, he had already been using the terminal since 1971.

McCarthy's papers, but also the remnants of his material practices of personal recordkeeping, writing, documenting, and cataloging[345].

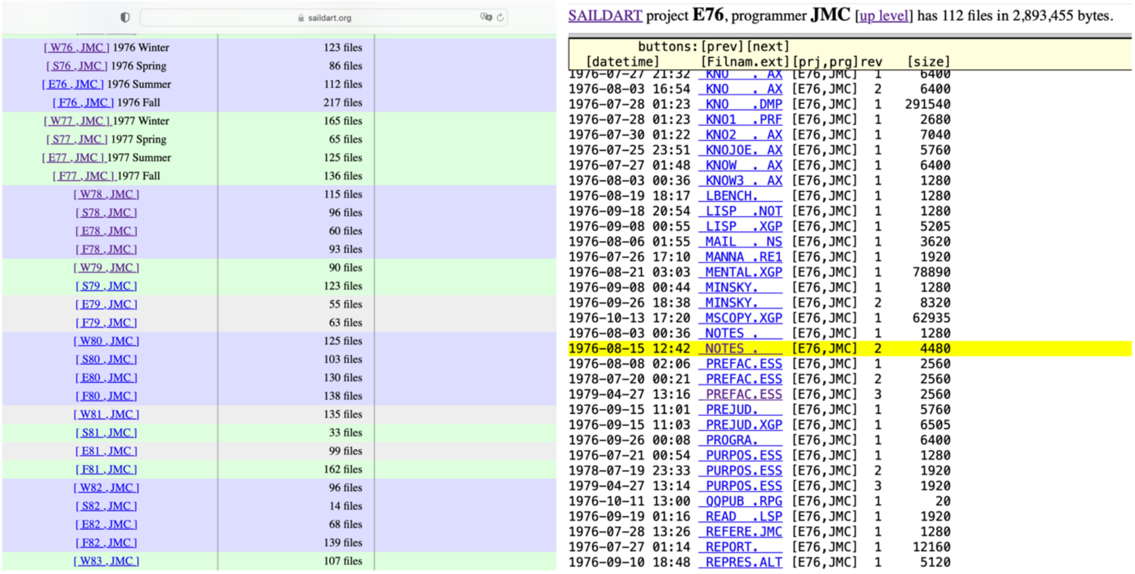
Surprisingly enough, McCarthy's directory represents, above all, an impressive laboratory of writing. Describing his activities at SAIL in one of the oral histories, McCarthy noted: "my own use of the computer was not CPU-intensive. I never did use it for anything but typewriter, for writing papers"[346]. Quite surprisingly, the McCarthy archive indeed resembles more of a conventional humanities archive than a software repository. At a glance, the archive offers a simple answer to the question of what McCarthy was engaged in at SAIL: McCarthy was writing. Not programs, but texts. From 1972 to 1990, he kept nearly 15,000 files, the bulk of which were essays and memoranda, but also literary sketches (including hobbit biographies), outlines of articles and reviews, short stories, riddles, parabolas, thought experiments, his own aphorisms, reflections on conversations and encounters with people, quotation lists for future reference.

SAILDART not only discloses an impressive number of texts written by McCarthy, but also reveals that writing over the years had been integrated into his everyday practices. Since 1975, McCarthy's directory was arranged chronologically: each season of the year had its own folder (Fig. 4.12). On average, each folder contained fifty to a hundred files. Each season, McCarthy worked on a number of subjects and topics, some of which carried over into the next "season."

These "seasonal" directories bring into focus the temporality or duration of inventions. What in the biography appears as a singular event – for example, the invention of the method of circumscription – in the archive turns out to be the work of several seasons, during which ideas, drafts, bibliographies are accumulated. The archive shows how an idea takes shape across a multitude of documents and remnants, alongside a multitude of other projects, thoughts, sketches, before crystallizing into a publication.

345 The view of the researcher's archive as a testimony to his or her specific material practices is offered by Jean-François Bert (2014).

346 John McCarthy: An Interview Conducted by Peter Asaro with Selma Selma Šabanović, IEEE History Center, 8 June 2011 [https://ethw.org/Oral-History:John_McCarthy]



saildart.org

[W76 , JMC] 1976 Winter	123 files	
[S76 , JMC] 1976 Spring	86 files	
[E76 , JMC] 1976 Summer	112 files	
[F76 , JMC] 1976 Fall	217 files	
[W77 , JMC] 1977 Winter	165 files	
[S77 , JMC] 1977 Spring	65 files	
[E77 , JMC] 1977 Summer	125 files	
[F77 , JMC] 1977 Fall	136 files	
[W78 , JMC]	115 files	
[S78 , JMC]	96 files	
[E78 , JMC]	60 files	
[F78 , JMC]	93 files	
[W79 , JMC]	90 files	
[S79 , JMC]	123 files	
[E79 , JMC]	55 files	
[F79 , JMC]	63 files	
[W80 , JMC]	125 files	
[S80 , JMC]	103 files	
[E80 , JMC]	130 files	
[F80 , JMC]	138 files	
[W81 , JMC]	135 files	
[S81 , JMC]	33 files	
[E81 , JMC]	99 files	
[F81 , JMC]	162 files	
[W82 , JMC]	96 files	
[S82 , JMC]	14 files	
[E82 , JMC]	68 files	
[F82 , JMC]	139 files	
[W83 , JMC]	107 files	

SAILDART project **E76**, programmer **JMC** [up level] has 112 files in 2,893,455 bytes.

buttons: [prev] [next]

[datetime]	[Filnam.ext]	[prj,prg]	rev	[size]
1976-07-27 21:32	KNO . AX	[E76,JMC]	1	6400
1976-08-03 16:54	KNO . AX	[E76,JMC]	2	6400
1976-07-28 01:23	KNO .DMP	[E76,JMC]	1	291540
1976-07-28 01:23	KNO1 .PRF	[E76,JMC]	1	2680
1976-07-30 01:22	KNO2 . AX	[E76,JMC]	1	7040
1976-07-25 23:51	KNOJOE. AX	[E76,JMC]	1	5760
1976-07-27 01:48	KNOW . AX	[E76,JMC]	1	6400
1976-08-03 00:36	KNOW3 . AX	[E76,JMC]	1	1280
1976-08-19 18:17	LBENCH.	[E76,JMC]	1	1280
1976-09-18 20:54	LISP .NOT	[E76,JMC]	1	1280
1976-09-08 00:55	LISP .XGP	[E76,JMC]	1	5205
1976-08-06 01:55	MAIL . NS	[E76,JMC]	1	3620
1976-07-26 17:10	MANNA .RE1	[E76,JMC]	1	1920
1976-08-21 03:03	MENTAL.XGP	[E76,JMC]	1	78890
1976-09-08 00:44	MINSKY.	[E76,JMC]	1	1280
1976-09-26 18:38	MINSKY.	[E76,JMC]	2	8320
1976-10-13 17:20	MSCOPY.XGP	[E76,JMC]	1	62935
1976-08-03 00:36	NOTES .	[E76,JMC]	1	1280
1976-08-15 12:42	NOTES .	[E76,JMC]	2	4480
1976-08-08 02:06	PREFAC.ESS	[E76,JMC]	1	2560
1978-07-20 00:21	PREFAC.ESS	[E76,JMC]	2	2560
1979-04-27 13:16	PREFAC.ESS	[E76,JMC]	3	2560
1976-09-15 11:01	PREJUD.	[E76,JMC]	1	5760
1976-09-15 11:03	PREJUD.XGP	[E76,JMC]	1	6505
1976-09-26 00:08	PROGRA.	[E76,JMC]	1	6400
1976-07-21 00:54	PURPOS.ESS	[E76,JMC]	1	1280
1978-07-19 23:33	PURPOS.ESS	[E76,JMC]	2	1920
1979-04-27 13:14	PURPOS.ESS	[E76,JMC]	3	1920
1976-10-11 13:00	QQPUB .RPG	[E76,JMC]	1	20
1976-09-19 01:16	READ .LSP	[E76,JMC]	1	1920
1976-07-28 13:26	REFERE.JMC	[E76,JMC]	1	1280
1976-07-27 01:14	REPORT.	[E76,JMC]	1	12160
1976-09-10 18:48	REPRES.ALT	[E76,JMC]	1	5120

Figure 4.12. On the left is a snippet of the JMC directory structure, on the right is an example of a folder for the summer of 1976

All of these many ideas, files, and drafts were digitally inventoried. Each season, McCarthy made an index of all the (important) files, with some annotations (Fig. 4.13), as well as produced an automatic document indicating when these files had been dumped. In fact, documentary and mnemonic practices formed the object of his ongoing reflection: he discussed the problem of organizing files[347], outlined a digital library project[348], reviewed SAIL experience of using computers in the office[349].

On its own, this grandiose scriptural economy is nothing short of remarkable. Even by the standards of the AI community, McCarthy published very little: unlike the other "fathers of AI" – Herbert Simon, Allen Newell, or Marvin Minsky – he did not issue a single monograph, and most of his publications were barely more than three pages long. McCarthy's publication history, in other words, in no way reveals his engagement with writing. The archive, on the other hand, uncovers the traces of his *digital* scriptural economy: rather than publishing his writings, McCarthy communicated them via computer (or more precisely, the information terminal).

347 https://www.saildart.org/FILES.PRO[S79,JMC]1
348 https://www.saildart.org/LIBRAR[S79,JMC]2
349 https://www.saildart.org/OFFICE.2[W80,JMC]

```
BUSH.NS[W80,JMC]        Bush's standard speech
CAMPBE.[W80,JMC]        Notes on %2Descriptive Epistemology: Psychological,
                        Sociological, and Evolutionary%1
                        by Donald T. Campbell
CAUSE.[W80,JMC]         CAUSALITY IN SYSTEMS OF INTERACTING AUTOMATA
CHALLE.[W80,JMC]        Challenges to weak AI, especially to Schankian AI
CIRCUM.REP[W80,JMC]     REPLIES TO COMMENTS ON "FORMALIZING OCKHAM'S RAZOR"
CIRCUM.W80[W80,JMC]     Circumscription paper
CONAES.[W80,JMC]        Questions on the CONAES report
CONNAL.NS[W80,JMC]      Connally's standard speech
COUNTE.[W80,JMC]        CARTESIAN COUNTERFACTUALS
CREARY.[W80,JMC]        conversations with Creary
DISPLA.[W80,JMC]        display supply conversations
DISSEN.[W80,JMC]        Dissenting opinions (mine)
DOYLE.[W80,JMC]         comments on Doyle's Stanford lecture
ELEPH2.[W80,JMC]        elephant notes
ENERGY.BIO[W80,JMC]     biography for energy lectures
ERROR.[W80,JMC]         some common misused words
FEYERA.[W80,JMC]        for the philosophical lexicon
FRAME.[W80,JMC]         THE FRAME PROBLEM IN THE 1980s
HOOK.[W80,JMC]          A SKYHOOK SUPPORTED BY ORBITING PARTICLES
HOOK.M[W80,JMC]         A SKYHOOK SUPPORTED BY ORBITING PARTICLES
IDEAS.W80[W80,JMC]      Scenes a la Barwise
                        Advice Taker should begin with declarative formalism
                        AIphil discussion on individual ascriptions
                                 of mental qualities
                        anthroplogists overaccept primitive ideas
                        differential equations
                        Circumscription and intensionality
                        Going down
ISRAEL.SPE[W80,JMC]     Speech at IBM symposium in Haifa
JOINT.[W80,JMC]         JOINT KNOWLEDGE
KENNED.NS[W80,JMC]      Kennedy's standard speech
LOSER.[W80,JMC]         IRS demands that banks keep records IRS loses
LYSENK.[W80,JMC]        Lysenkoism in Energy Research
META.[W80,JMC]          Metaphilosophy
MINSKY.SE1[W80,JMC]     Minsky on Searle
NAMING.[W80,JMC]        AI and "Naming, Necessity and Natural Kinds"
OFFICE.[W80,JMC]        CRITERIA FOR USEFULNESS OF COMPUTERS IN OFFICES
OFFICE.2[W80,JMC]       CRITERIA FOR USEFULNESS OF COMPUTERS IN OFFICES
OPPORT.[W80,JMC]        Social effects of keeping publications in the computer
POSH.NS[W80,JMC]        top hotels and restaurants
PROBLE.[W80,JMC]        Bill wants to have an affair with Tom's wife
REAGAN.NS[W80,JMC]      Reagan's standard speech
RENEW.PRO[W80,JMC]      renewal of nsf proposal
ROMANT.[W80,JMC]        Pro-technology must appeal to the imagination
SE2.[W80,JMC]           notes on SE2 affairs
SEARLE.[W80,JMC]        short and sweet Searle refutation
SEARLE.1[W80,JMC]       notes on Searle
SEARLE.2[W80,JMC]       Comments on Searle's "Minds, Brains and Programs"
SECRET.NS[W80,JMC]      News stories that wrongly reveal secrets
SEX.[W80,JMC]           Comments on Evolution of Human Sexuality
SOCIAL.[W80,JMC]        SOCIAL SCIENCE IS VERY COMPLICATED
STANDA.[W80,JMC]        TARGETS FOR STANDARDIZATION
TALK.[W80,JMC]          notes for CASBS on computerization
TALK1.[W80,JMC]         SOCIAL OPPORTUNITIES FROM COMPUTER TECHNOLOGY
TECSOC.[W80,JMC]        SOCIETAL ATTITUDES TO TECHNOLOGY - for ACM
```

Figure 4.13. McCarthy's inventory for winter 1980

The dominant genre in McCarthy's directory is the memorandum, a genre of informational writing *par excellence* (Guillory 2004). McCarthy wrote memos on literally everything: the range of topics extended from narrow technical and scientific issues – the modality in first-order logic[350] or the state-of-the-art in robotics[351] – to reflections on the idea of happiness[352] or on the mind in dogs[353], or on storing data in molecules[354]. Whether a reflection on happiness or robotics, McCarthy's memo was a very concise, succinct and clearly worded[355] text, often having an articulated structure: lists, formulas, numerations. It usually resembled a logical chain, a long syllogism and reminds of the structures of list processing language (LISP) invented by McCarthy. The style of McCarthy's memos clearly expressed and anticipated the

350 https://www.saildart.org/MODAL[S78,JMC]

351 https://www.saildart.org/QUASAR[S78,JMC]

352 https://www.saildart.org/HAPPIN[E76,JMC]

353 https://www.saildart.org/DOGMIN[E77,JMC]2

354 https://www.saildart.org/MEMORY[1,JMC]. Today the technique McCarthy describes is known as molecular memory. I have not been able to find its origins or any discussion of it before the 2000s. McCarthy forwarded his thoughts on memory to Licklider and asked him for funding for a chemist: https://www.saildart.org/LICK[ESS,JMC].

355 For a discussion of how these principles are incorporated into the history of the genre, see Guillory 2004.

new economy of attention, characteristic of the new digital practices of reading and dealing with information.

Joanne Yates (1989) locates the emergence of memorandum in the late 19th and early 20th centuries and relates it firstly to the rise of new managerial practices and secondly to new technologies of writing and storage (primarily the advent of typewriters and vertical files). In McCarthy's case, the memo was also inextricably linked to the evolving digital technologies of writing, storage, and communication. A memo was that which was written on a computer[356] (or a "home information terminal") and that which was disseminated by means of computer networks. It was a "note to oneself" (Guillory 2004, 116) shared through networks – first the narrow community of SAIL and ARPANET, then USENET, and finally, with the advent of the WEB, through his personal websites.

The remnants of the memo's technological background may be located in the archive. McCarthy's directory has a special file that set the rules for formatting memorandums, as well as a custom file that automatically provided memos with a header and signature[357]. The latter included McCarthy's ARPANET account and the date and time when the draft was "PUBbed" – formatted for on-screen display or print[358]. "PUBbing" signaled that the text is not just a working draft hidden from public view. It suggested that the text was meant to be read by other users of the system; that it was available through the very "information terminals" as McCarthy described in Bordeaux. At one point, in 1976, McCarthy even named one of his directories "McCarthy's electronic magazine", with the idea of publishing his texts and responses to them in this separate directory[359].

What the researcher finds today in McCarthy's directory is not drafts and sketches for his internal use, which would subsequently be crystallized into a publication, a result visible to the public. This configuration would be typical for "traditional" collections of personal papers: "personal" is that which is not visible to the public, that which remains behind the scenes in the communication of scientific achievements. Instead, McCarthy's directory constitutes an *information space* which draws no boundaries between (scientific) activity and its communication. McCarthy's files were instantly made public, communicable, sharable in the

356 From the many drafts, excerpts, and sketches kept in the SAILDART, one to ascertain that McCarthy was writing memos directly on his computer.

357 https://www.saildart.org/MEMO[LET,JMC]1. Formatting files have changed over the years; all MEMO files can be found in the [LET, JMC] directory.

358 For a discussion of PUB, a document compiler used at SAIL in the 1970s, see the next chapter.

359 https://www.saildart.org/CONTEN[PUB,JMC]2

AI circle (or a circle united by ARPANET), open for discussion and response. It is this information space that is made available to us through the SAILDART.

Information terminals were giving birth to new science communication practices and blurring the distinction between the private, enclosed cabinet or laboratory in which science is done and the communication of science to the public. These new kinds of scientific communication, all that was not published but communicated through the computer, however, remain out of the bounds of scientific biography and its conceptions of "intellectual events".

4.4.3. Defending AI and philosophizing AI

Having outlined McCarthy's scriptural practices, I now turn to his literary and social technologies in advocating, defending, and contextualizing the discipline of Artificial Intelligence.

McCarthy was involved in *every* controversy surrounding AI, reviewing, responding to, participating in, and initiating debates. He reacted to James Lighthill's critique[360], reviewed Joseph Weizenbaum's book on AI ethics[361], wrote responses to John Searle[362] and Hubert Dreyfus[363], who called into question the intelligence of the machine[364]. Moreover, in a few of these cases, he himself initiated public debates with his opponents: in the archive one can find his calls for debate to Weizenbaum[365] and Searle[366].

The responses and reviews, made in reaction to "various attacks on AI" (McCarthy, 1996, vii), made up McCarthy's one and only small book, published in 1996. More of a "grey paper" than a full-fledged collection, it was composed of short texts written in the same "memo style": very concise, with numbering lists and bullet points[367]. The volume was prefaced by a short introduction from the mid-1990s, in which McCarthy summarized the following major lines of attack against AI:

---

360 See, for example, https://www.saildart.org/LIGHT.2[2,JMC]

361 https://www.saildart.org/WEIZEN[W76,JMC]

362 See, for example, https://www.saildart.org/SEARLE.RE1[E79,JMC]2

363 https://www.saildart.org/DREYFU.REV[F78,JMC

364 Since in this case I am focused on the McCarthy's collection and its affordances, I will not present the course and content of the controversies. They are all well-known and thoroughly documented: cf. McCorduck 1978, Gardner 1985, Boden 2006.

365 https://www.saildart.org/CHALLE[S77,JMC]

366 https://www.saildart.org/SEARLE[E84,JMC]. McCarthy's debate with Lighthill transmitted by the BBC is also well-known.

367 The essays were published exactly as they stand in the archive, without further editing. Holding the printed version, one can distinctly feel the gap between the text made for print and the one made for digital communication.

- AI as "incoherent concept philosophically" (Dreyfus, Searle)

- AI being "immoral" (Weizenbaum)

- AI being mathematically impossible (Penrose)

- AI not making progress (Lighthill + Dreyfus)

Each of the "AI opponents" received about a dozen files within McCarthy's directory. McCarthy corresponded with some of them, outlined arguments and defense strategies, wrote reviews and responses (Fig. 4.14). Some of the "debates" had lasted for decades. For instance, McCarthy revisited the Lighthill report criticizing AI in the 1970s twenty years later[368], and his correspondence with Searle had spanned more than a dozen years[369]. The archive thus makes it possible to recognize that *defending AI* for McCarthy was by no means a side activity.

1973-06-14 22:04	LIGHT .	[2,JMC]	1	1805
1973-06-14 23:58	LIGHT . 2	[2,JMC]	1	3840
1973-06-17 01:45	LIGHT . 3	[2,JMC]	1	2105
1973-06-22 22:30	LIGHT . 4	[2,JMC]	1	4480
1973-06-22 23:07	LIGHT . 5	[2,JMC]	1	1280
1973-06-23 16:48	LIGHT . 6	[ESS,JMC]	1	2560
1974-03-13 10:07	LIGHT .COM	[2,JMC]	1	16000
1974-04-11 14:27	LIGHT .DOC	[2,JMC]	1	24965
1973-12-19 16:48	LIGHT .RE2	[2,JMC]	1	5760
1974-03-11 14:48	LIGHT .RE3	[2,JMC]	1	14720
1974-03-21 20:28	LIGHT .RE3	[2,JMC]	2	18275
1974-04-08 21:17	LIGHT .RE3	[2,JMC]	3	15360
1974-03-22 18:33	LIGHT .RE4	[2,JMC]	1	1280
1974-04-08 21:14	LIGHT .RE4	[2,JMC]	2	5760
1974-04-11 12:19	LIGHT .RE5	[2,JMC]	1	19200
1974-04-22 15:41	LIGHT .RE5	[2,JMC]	2	19200
1974-06-18 23:01	LIGHT .RE5	[2,JMC]	3	22605
1974-10-02 02:42	LIGHT .RE5	[ESS,JMC]	1	22605
1975-07-25 14:22	LIGHT .RE5	[ESS,JMC]	2	19200
1980-02-21 18:13	LIGHT .RE5	[ESS,JMC]	3	19200
1973-10-26 01:46	LIGHT .REV	[2,JMC]	1	1610
1973-11-10 17:28	LIGHT .REV	[2,JMC]	2	3200
1973-12-19 00:24	LIGHT .REV	[2,JMC]	3	4480
1973-11-10 18:21	LIGHT2.REV	[2,JMC]	1	8960
1990-03-09 15:16	LIGHTH.	[W90,JMC]	1	7680
1990-03-23 15:57	LIGHTH.	[W90,JMC]	2	7680
1990-03-03 23:10	LIGHTH.LOG	[1,JMC]	1	320
1990-03-09 14:42	LIGHTH.LOG	[1,JMC]	2	325
1990-03-23 15:57	LIGHTH.LOG	[1,JMC]	3	325
1990-03-09 15:16	LIGHTH.LOG	[W90,JMC]	1	315
1973-10-08 01:55	LIMITS.ESS	[ESS,JMC]	1	3840
1985-05-31 15:30	LIN . 1	[LET,JMC]	1	5760
1987-01-13 12:45	LIN . 2	[LET,JMC]	1	1280
1979-03-01 12:14	LINDZE. 1	[LET,JMC]	1	3200

Figure 4.14. Files on James Lighthill in McCarthy's directory.

Prior to being published McCarthy's reviews and responses became the subject of peer discussion "online". McCarthy shared the file with colleagues while it was still in the writing stage so that they could revise it, comment on it, contribute to it. So, for example, in a message to Marvin Minsky of MIT, McCarthy specified a file with his notes for a future review of Weizenbaum's book, and, also asked him some factual questions needed for his argument:

368 https://www.saildart.org/LIGHTH[W90,JMC]

369 Correspondence with Searle extended up to 1990 (as the archive does), very probably for longer. McCarthy's last mail to Searle, as recorded in the archive is https://www.saildart.org/SEARLE.1[LET,JMC].

```
∂16-MAR-76 2316 JMC
To: minsky @ MIT-AI
WEIZEN[W76,JMC] contains notes for a review. When it is finished, it will be WEIZEN.REV[PUB,JMC]. There are two quotes in the book whose authors are withheld. One of them is almost certainly Fredkin, and I'll check it with him, but the other I am not sure about but suspect it might be Warren McCullogh. Here it is, and if you have any way of checking, I would be thankful[370].
```

The following message – apparently an excerpt from correspondence with Randall Davis (KRD) – makes it explicit that members of the AI community also proofread and commented on the review prior to its publication:

```
22-MAR-76       JMC
To: KRD
Thanks for the corrections and the comments. You are right that there is too to be discussed for a review. Your guess that Weizenbaum "makes predictions" is mistaken. He says less about it than I did on that one page[371].
```

Another example is the notes file for the upcoming debate with Searle on the Chinese Room experiment. McCarthy's notes show evidence of discussions with colleagues giving him pieces of advice on how to proceed and what line of argument to take during the discussion:

```
searle[e84,jmc]         Notes for 1984 Sept 4 discussion with John Searle

Suppes's advice
Ask for details. He's strong on generalities.
He's not strong on logic.
Stay away from Wittgenstein's objections to private language.
There's Kripke's new book, but you won't read that by Tuesday.
In general the 1st vs. 3rd person view hasn't been formulated[372].
```

Defending AI from external attacks thus appears as both a personal and collective project. McCarthy's reviews were read and commented on by his colleagues, and he in turn read and commented on their texts (McCarthy's directory, for example, also held Minsky's essay in response to the polemic with Searle[373]). In the discussion of both the "attacks" and the strategies for responding to them, a space of commonality and collective agreement was formed. Although within AI its different representatives may have disagreed with each other – for example, McCarthy could quite openly criticize Minsky's theory of frames (Hilts 1984), – in responding to the critics they put forward a unified, shared, concerted standpoint. The overall intent behind this move was to limit the outside judgments of AI while giving a voice to AI

370 https://www.saildart.org/OUTGO.MSG[ESS,JMC]7
371 Ibidem.
372 https://www.saildart.org/SEARLE[E84,JMC]
373 https://www.saildart.org/MINSKY.SE1[W80,JMC]

practitioners. To put it differently, it was only the AI community itself that should be passing judgment on the discipline.

This attitude is most notably evident in McCarthy's relationship with the humanities and social sciences. Perhaps more than any other discipline, AI has become the scene of a clash between science and the humanities. The field of Artificial Intelligence attracted great interest from philosophers, sociologists, anthropologists, psychologists[374] and sparked numerous debates about what were considered to be *their* objects of study: mind, knowledge, intentionality, or human-machine interaction.

It is the attitudes of the humanities and philosophy, their drive to pass judgment on AI and evaluate AI's potential, that McCarthy most fiercely opposed. Consider, by way of example, his review of "The Question of Artificial Intelligence" – one of the early ventures to describe the AI field from a sociological perspective in no way intended as an *attack* on AI. Written in his signature and inimitable manner, McCarthy's review opened as follows:

```
This book is of a genre that treats a scientific field using various social
science and humanistic disciplines,e.g. philosophy, history, sociology,
psychology and politics. Scientists often complain about the results, both
generally (judging the whole effort as wasted) and specifically (citing
instances of ignorance and misunderstanding). I'm open minded about the
general activity; maybe the sociology of research in AI has independent
intellectual interest, though surely less than that of AI itself.
[…]
 This review mainly concerns specific matters, and is mainly negative,
complaining about ignorance and prejudice. The review also contains
some suggestions about how this kind of thing can be done better ---
assuming it is to be done at all.[375]
```

Commenting on each of the collection's articles in turn, McCarthy put forward the following lines of critique. First, the authors cited the sociological literature, rather than the AI literature (at this point McCarthy listed specific titles that should have been referenced: "articles in 'Artificial Intelligence', the proceedings of the International Joint Conferences of AI"). Second, in describing the AI community, the authors were not talking to AI researchers themselves. Third, the historical accounts of AI were inaccurate and even fabricated (at this point he invoked his authority as a first-hand participant in that history).

Ultimately, in McCarthy's perspective, "the ignorance and prejudice" of the authors stemmed from the fact that they did not give voice to the AI community, instead imposing their external view and evaluation on the field. In the end, the narratives offered by scholars in the

374 A more in-depth discussion of this interest can be found in the next chapter.
375 https://www.saildart.org/REVIEW[W88,JMC]4

humanities or social scientists were taken as an intervention into the AI realm by "non-specialists" who were unable to judge what was going on within the discipline and therefore, in McCarthy's words, had to "invent the subject". Even the AI history thus very early on came to be a battleground on which the right to tell it and to judge it were fought over. For a good reason, McCarthy wrote (and, remarkably, published!) a number of articles on the history of his own inventions, including that of LISP (McCarthy 1978) or Nonmonotonic reasoning (McCarthy 2004).

McCarthy, however, was not just *defending* the boundaries of AI from the intrusion of social researchers. In his writings he himself was quite self-consciously stepping into the domain of the humanities and social sciences. What McCarthy's archive reveals is precisely his breadth of the humanist and social reflection. In one of his half-written essays on politics, for example, he remarked:

```
These essays are mainly concerned with ways of improving human life by introducing new technology.
[…]
Perhaps I just haven't read the right literature, but I have not been able to find a discussion of this point by certified social scientists, so this is strictly an amateur effort to analyze why this is so and what might be done about it.[376]
```

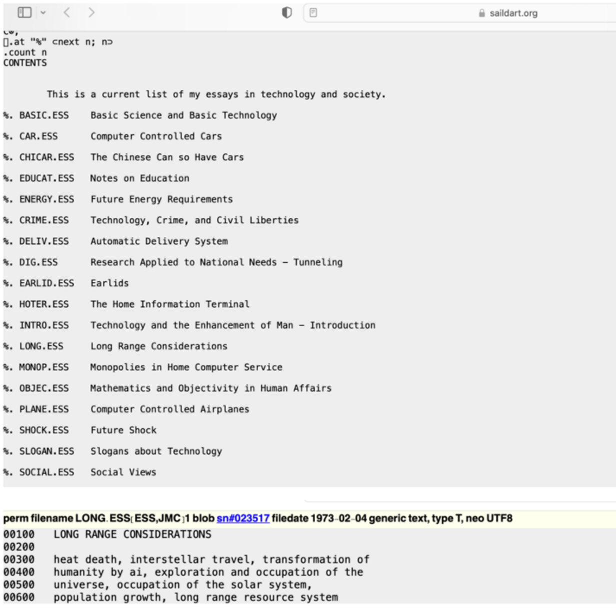
The main focus of this reflection could be labeled as "technology and the future of man". McCarthy wrote dozens of essays and notes on technology and society that fold into a coherent political, technical, social program encompassing both "long range considerations"[377] ("interstellar travel", "transformation of humanity by AI", "occupation of the universe") and, for example, the visionary "earlid" proposal (Fig. 4.15). This was a techno-optimist project, in which technologies were given power to address societal problems ranging from women's liberation[378] to controlling airplane traffic[379].

[376] https://www.saildart.org/POLIT.ESS[ESS,JMC]1

[377] https://www.saildart.org/LONG.ESS%5bESS,JMC%5d1; For fuller version see https://www.saildart.org/LONG.ESS[ESS,JMC]3

[378] https://www.saildart.org/WOMEN.ESS[ESS,JMC]

[379] https://www.saildart.org/PLANE.ESS[ESS,JMC]



saildart.org

.at "%" ⊂next n; n⊃
.count n
CONTENTS

This is a current list of my essays in technology and society.

%. BASIC.ESS Basic Science and Basic Technology
%. CAR.ESS Computer Controlled Cars
%. CHICAR.ESS The Chinese Can so Have Cars
%. EDUCAT.ESS Notes on Education
%. ENERGY.ESS Future Energy Requirements
%. CRIME.ESS Technology, Crime, and Civil Liberties
%. DELIV.ESS Automatic Delivery System
%. DIG.ESS Research Applied to National Needs – Tunneling
%. EARLID.ESS Earlids
%. HOTER.ESS The Home Information Terminal
%. INTRO.ESS Technology and the Enhancement of Man – Introduction
%. LONG.ESS Long Range Considerations
%. MONOP.ESS Monopolies in Home Computer Service
%. OBJEC.ESS Mathematics and Objectivity in Human Affairs
%. PLANE.ESS Computer Controlled Airplanes
%. SHOCK.ESS Future Shock
%. SLOGAN.ESS Slogans about Technology
%. SOCIAL.ESS Social Views

perm filename LONG.ESS[ESS,JMC]1 blob sn#023517 filedate 1973-02-04 generic text, type T, neo UTF8

LONG RANGE CONSIDERATIONS

heat death, interstellar travel, transformation of
humanity by ai, exploration and occupation of the
universe, occupation of the solar system,
population growth, long range resource system

Figure 4.15. Top is a list of McCarthy's 1973 essays, bottom is the outline of the "Long Range Considerations".

In thinking about the social, McCarthy's method was often a thought experiment in which the "right solution" came from technology. A typical example was the "Doctor's dilemma"[380] telling of a doctor who was able to cure any disease by simply touching the patient. What would happen to such a doctor? How could he bring maximum benefit to mankind? How to choose whom to cure in the limited span of the doctor's life? After listing all imaginable scenarios ("exercises in pessimism and paranoia") from burning the doctor at the stake by religious fanatics to a nuclear war with Russians for the right to own his gift, McCarthy cited "the solution from common sense and technology". Based on calculations, it went as follows: a doctor would actually be able to cure all the people in the world whose illness would allow them to get to the doctor on time. The solution was grounded in mortality statistics, birth rates and included a proposal to build special transportation systems in order to reduce mobility costs. To borrow one of McCarthy's own aphorisms, the moral of the matter was that

`The best way to solve a moral problem is to turn it into a technical problem.`[381]

McCarthy's essays and memos formed a coherent techno-social utopia in which technology took the place of Baconian knowledge: "*technology* is power". Among the envisioned technologies, AI, of course, occupied a prominent place. McCarthy did not so much discuss artificial intelligence *per se*, but rather the ways in which it could be integrated into

380 https://www.saildart.org/DOCDIL.LIT[ESS,JMC]1
381 https://www.saildart.org/SLOGAN.ESS[ESS,JMC]2, slogan 19.

life, politics, and society. He examined, for example, the advantages of computer-controlled cars[382] or the enhancement of the individual through computer-aided design systems[383].

This program is certainly awaiting its researcher, for it can contribute not only to McCarthy's scientific trajectory, but also to the (history of) reflection about Artificial Intelligence and the ever-evolving debates about the place AI should take in social relations, in the labor markets, politics, education, or culture. McCarthy's project could also be of interest in terms of the institutionalization and self-consciousness of the discipline of AI. As we have seen, AI was in need not only of "scientific breakthroughs" – discoveries and inventions – but also of this kind of reflection outlining how and why it was important to the "world" and to the future.

As for McCarthy's scientific persona, his collection allows one to discern a much more many-sided personality of McCarthy, and beyond his lists of publications and inventions, to see a much more expansive AI program. The archive allows one to trace the dynamics of McCarthy's views as well as the way he put them into action: the directory contains McCarthy's letter to Richard Nixon, many traces of McCarthy's interaction with the Soviet scientists, his petitions, letters to journals, and so on.

Both defending and philosophizing AI fall outside the conventional image of the AI scientist and are not included in any of McCarthy's biographies (it probably wouldn't have been missed in a biography of a humanist, where the facts of invention are remarkably absent). The archive thus *deconstructs* the scientific biography, shifting the focus from what is considered important to the AI scientist – inventions, publications, awards – to a different part of the scientific life and scientific identity.

382 https://www.saildart.org/CAR.ESS[ESS,JMC]3
383 https://www.saildart.org/ENHANC.ESS[ESS,JMC]

4.5. Residual commons: from discourses to practices

Whereas the previous chapter looked at the archive as a record of a personal trajectory, this one takes a glimpse at how the archive reflects the *communal*. The focus of this chapter revolves around the question of scientific community and social technologies at SAIL. What does the archive bring to or shift in our views of the SAIL community? What can it possibly uncover about its identity, self-definition and self-understanding, about what makes members of the laboratory a "thought collective"?

4.5.1. Wizards and positivists

As in the preceding chapter, I will begin with a brief overview of the different accounts of SAIL's communality. In very rough strokes, I will describe the two main ways of describing the AI community as being in tension with each other. This overview in no way pretends to be comprehensive and certainly oversimplifies the story.

The first type of narrative represents a view of the SAIL community as a subculture. This mode of representing and (self-) understanding of the SAIL community began with the famous article by Stewart Brand (1972), who arrived at SAIL in the middle of the night in 1972 to find a crowd of young "long-haired technicians" playing the game Spacewar. Editor of the iconic counter-culture *Whole Earth Catalog*, Brand described what he saw exactly in terms of counter-culture, as a geeky, alternative way of life of the young people who were fascinated with technology, opposed to the Vietnam War and conventional hierarchies, who did not separate work and leisure, sported weird hairstyles, often had no degree, but nevertheless became "computer wizards".

Brand's view of SAIL as a subculture has been adopted by other prominent AI journalists. For example, Steven Levy, in his famous book "Hackers", drew comparisons between the MIT and SAIL communities as embodiments of East and West Coast cultures (Levy 2010 [1984], 134-139). John Markoff likewise described SAIL as a "self-possessed subculture" and a "hacker's paradise but far different from the engineering-centric world of MIT" (Markoff 2006, 95).

This view of the uniqueness of the socio-cultural environment and atmosphere at SAIL is echoed in the memoirs of the SAIL community. Les Earnest, in his memoir "SAIL Away"[384], talks about the "SAIL culture", which, for example, some employees had taken to Xerox PARC. Bruce Buchanan, in his introduction to the SAIL's memos edition, describes the "near

384 https://web.stanford.edu/~learnest/sail/sailaway.htm

magical qualities of the SAIL atmosphere", its "storybook flavor" and "sense of living in a fantasy world" (Buchanan 1983, 41). In reinforcing this image of a fantasy universe, the authors usually cite a shared collective passion for science fiction and video games, common fascination with technology[385], and some communal documents, e.g., the Jargon file[386], a vocabulary of hackers' slang, or the Yum-Yum file[387], a collaborative SAIL review of restaurants and cafes (Fig. 4.16).

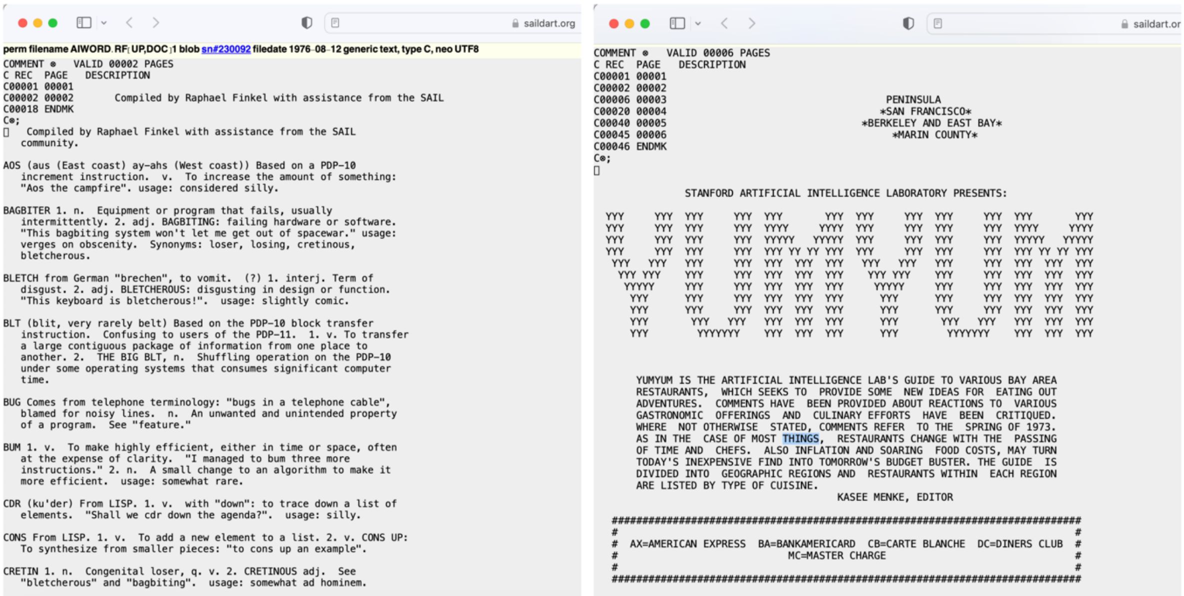


Figure 4.16. On the right is a fragment of the "Jargon file", on the right is a fragment of "YumYum".

It is in this playful, "techno-magical" and visionary ecosystem that both the community members and their storytellers see the origins of the lab's innovations. As Markoff puts it,

> SAIL was as unconventional as it was innovative. Researchers lived in the attic above their offices, encounter groups met in the steam tunnels in the basement, and from that tumult emerged the technological insights that would help reshape both Silicon Valley and the entire world during the next decade (Markoff 2006, xxii).[388]

In such accounts, the "scientific" history of SAIL appears as a natural continuation of this playful culture; "innovation" is thought of as a manifestation of freedom, youth, and unconventionality.

---

385 An analytical account of this discourse was presented by Paul Edwards, describing it as a "cyborg discourse" (Edwards 1997).

386 One of the early versions of the file in SAILDART: https://www.saildart.org/AIWORD.RF[UP,DOC]1

387 One of the early versions of the file in SAILDART: https://www.saildart.org/YUMYUM[P,DOC]1.

388 As Les Earnest puts it in"SAIL Away", "In SAIL's enjoyable work environment, researchers did pioneering work on computer vision, robotics, and automated assembly as well as mathematical theory of computation, theorem proving, and "common sense" reasoning". [https://web.stanford.edu/~learnest/sail/sailaway.htm]

The antithesis to the nostalgic and subcultural narratives about the AI community comes from the social science researchers: anthropologists, psychologists, and sociologists[389]. AI laboratories became a focus of anthropological and sociological interest quite early on. In 1984, Sherry Turkle issued her *The Second Self* based on observations at the MIT AI lab; in a 1985 article Steve Woolgar called for a study of "sociology of intelligent machines"; in 1987 Lucy Suchman published her famous work on human-machine interaction based on her work at Xerox PARC; by the turn of the 1980s-1990s Diana Forsythe published a series of articles concluding on community observation at Stanford. All of these works became classics of the social studies of AI (Turkle 2005 [1984], Woolgar 1985, Suchman 1987, Forsythe 2001)[390].

In opposition to the "techno-magic", social scientists put the mechanistic view and positivism at the heart of the community's world picture. According to Sherry Turkle, AI culture was organized around the idea of programming, which was claimed to be a universal interpretative framework (Turkle 2005 [1984], 222-223). Everything was explained through the prism of the program, including the human mind. Moreover, as Turkle shows, hackers and AI scientists[391] defined themselves through machines as well. Brian P. Bloomfield, describing the specificity of the AI community as a thought collective, in turn, emphasizes, above all, its "mechanistic view of knowledge" (Bloomfield 1987b, 82) and "technological determinism" (ibid., 85). Diana Forsythe in a number of articles also places scientism at the core of the AI community's identity:

> in the three-plus decades of the field's formal existence, researchers in AI have constructed a collective identity for themselves as practitioners of a "hard" science in which complex computer programs are used to illuminate experimentally the nature of intelligence (Forsythe, 2001, 82).

As Forsythe successively shows, in its self-understanding and its attitude toward work and knowledge, the AI community "deleted the social" and "deleted the cultural," instead opting for an oversimplified and mechanistic view (Forsythe 2001, 51-54).

As in the previous chapter, I will look at the outlines of collectivity in the archive not to challenge these accounts, but rather to question the SAILDART affordances. What structures of commonality can we discern through the archive? Does the archive complement, refine or elaborate on the accounts described above?

389 It is not for nothing that many of these narratives have been resisted by the AI community, as discussed in the McCarthy example.

390 Although the listed scholars did not conduct their research directly at SAIL, they draw conclusions about the AI culture and community at large, so I believe it is possible to extend their interpretations to the Stanford lab as well.

391 Turkle separates the identities of AI scientists and hackers. Some papers (see, for example, Bloomfield 1987) criticize this distinction.

In the following sub-chapters, I will move progressively toward deeper and deeper structures of collectivity, from communal discussions to shared knowledge to collaborative practices.

4.5.2. SOFT: the community speaks

In September 1974, Bruce Anderson, one of the lab programmers, initiated a discussion of the software goals at SAIL[392]. The discussion turned into a spontaneous "fileswap symposium": within two months he got responses from James McCarthy (JMC), Jeff Rubin (JBR), Terry Winograd (TW), Mitchell Model (MLM), Andy Moorer (music), Tom Knight (TK), Dave Smith (DAV), Brian McCune (BPM) and other members of the lab. In addition, in the wake of this debate, Terry Winograd launched a standalone discussion on personal computers.

Rather than being a closed private exchange of messages, the software goals debate took on the status of a forum open to all interested parties. All the replicas were collected in a file called "SOFT" and stored in a shared directory[393], deliberately displayed for public reading. Interestingly, half a century later, Baumgart also labels these archival files as being "insightful". In the logic of this chapter, this discussion appears to be just as insightful: through these files, we gain access into how the community spoke to itself back then, in the mid-1970s. SOFT represents the collective and spontaneous exchanges within the laboratory, not intended for an outside observer (e.g., an anthropologist). Neither, however, is it a nostalgic reminiscence of past times as they are viewed by lab members from the present day.

In his introductory remarks, Anderson raised the problem of the poor state of software in the laboratory. After long enumerating the shortcomings of various programs and utilities, he formulated a "diagnosis":

```
The basic problem is that the monitor and the programming languages used do
not impose the necessary common structure.
```

The solution he offered came down to developing a common uniform system, based on a single (brand new) programming language. In this system, the monitor and all system programs must have been compatible with this new language. As a final touch, he put this outlined "programming problem" at the center of all computer science:

```
Complete resolution of the programming problem for the community at large
is much of what Computer Science research is (or ought to be about).
```

392 https://www.saildart.org/SOFTWR.GOL[SOF,DOC]1

393 For a discussion of this directory – DOC – see the next paragraph.

The debate then took several directions. The first strand of the discussion, touched upon in one way or another by all participants, was the question of the effectiveness or functionality of the existing system. On this issue, stances differed considerably. A number of discussants took a critical stance: for example, McCarthy, who drew attention to the lack of a “subroutine library” at SAIL; or Mitchell Model, who criticized the system for the lack of interactive programs, the lack of standardization and many other issues. Others, on the contrary, mentioned “the advantages of a good document handling, distribution, and production system”, or even argued that SAIL “programs are pushing the state of the art in programming”.

Another line of discussion revolved around the idea of introducing a single standard programming language. On this point, as well, there was little agreement: so, one of the discussants (Dave Smith) explained in a very detailed response why LISP70 should have become a standard language of the lab, while others (Andy Moorer), on the contrary, rejected the very idea of a unified language, appealing to the specificity of their research and the heterogeneity of AI in general.

Further, a number of respondents questioned whether the issue was technical systems and instead claimed that the source of the problem was people. So, Jeff Rubin suggested that the problem lied more in the approach of the lab’s system programmers, who got so caught up in the idea of the program that lost sight of its users. Mitchell Model, on the other hand, saw the problem in the “lack of community” in general, in the deficiencies of exchange and communication between lab members.

For some contributors this discussion also provided an opportunity to outline their research and technical needs for their colleagues. This was the case of Terry Winograd, covering the specifics of research in natural language processing, Andy Moorer outlining the needs of digital music research and Brian McCune discussing the automatic programming group. Each of them described the technical part of his projects, indicated what systems or capacities he lacked, and described the contours of an imagined future technology that would help him carry out his research.

Finally, another stumbling point was the question of whether AI had a distinctive path compared to the computer science discipline. This issue was raised most explicitly in two final remarks: Tom Knight’s response (TK) and Anderson’s own postscript. The two took opposing positions: Knight defended the distinction between the two disciplines, while Anderson favored the homogenization of AI and computer science. Moreover, in Anderson’s view, solidarization with computer science was required for the AI project to succeed. As he concluded his postscript,

```
The field of AI has matured, and Stanford's position in it has changed. We
understand computers and computation much better, and we see that AI isn't
really so special and is very related to other disciplines. The days of the
computer-freak are over. It's time to try to state the purpose of the lab
and derive from them a plan for future computing facilities. If we fail to
do this, we may drift into a situation where we are constrained from
achieving many of our goals.
```

Even this very condensed recounting suffices to get a sense of how diverse were the stances, attitudes, and views taken by the SAIL programmers. The sought-after communality reveals itself not in the similarity of positions, but rather in the way in which these positions were articulated. What the discussion brings to light in this sense is that even as the lab members argued and described the dissimilarity of their professional trajectories, they were speaking one common, shared language. It was the language of software, systems, and technology by which both the future of the laboratory and current research were addressed.

Based on ethnographic observation, Diana Forsythe once noted that AI scientists tended to avoid defining their own identities or the scope of the AI field. As she pointed out,

> For example, asked to define 'artificial intelligence,' some knowledge engineers respond that they do not know what AI is. Rather than debate about the definition of "intelligence," or discuss whether AI is or not part of computer science, they prefer to get on with their systems (Forsythe 2001, 45).

At first glance, the software discussion is quite illustrative of Forsythe's observations: from the outset, Anderson quite consciously and explicitly abandoned the question of what AI was in favor of technical questions:

```
Aside: I could have taken a more thoroughgoing approach, and worked down
from "what is the Lab for?" or "what does ARPA think the Lab is for?" but
instead I went ahead and assumed that the Lab needs a software system which
enables people to write complex programs as easily as possible.
```

However, this very inclination to talk in terms of technologies does not mean that the researchers were not concerned with the boundaries of the discipline, or that they were insufficiently reflexive about their identity. The question of identity was simply addressed not by discussing the boundaries of concepts or methods, nor by questioning the limits of the Artificial Intelligence field – issues that are more of interest to social scientists. Instead, the collective identity was articulated in terms of technologies, systems, and programs. As we have seen, in this seemingly technical discussion about languages and monitors, SAIL programmers did address the future of the AI, they talked about community and their professional trajectories, they argued about how to evaluate the current state of AI research and whether AI forms part of computer science. But all these issues were resolved through the debates around technology: around what the systems of the future were, what the standardized AI language should have been, or whether the lab system should have supported a particular modification of LISP. This voice of the community, the way the community spoke about itself for itself and to itself, can only be heard through the archives.

4.5.3. DOC: the community shares knowledge

In discussing the mediology of the archive, I have already mentioned the time- and file-sharing system at SAIL, a system which incorporated and prescribed a certain notion of collectivity. Every file in any directory was available to all users; files were shared, exchanged, referred to, files were initially created with the view of the other. However, in addition to these sharing practices, SAIL had a well-established system of genres of communicating knowledge, findings, and inventions. In this part, I look at this system showing how the archive offers a glimpse into SAIL's knowledge sharing practices.

A synthesis of SAIL's documentation and genre system can be found in Les Earnest's 1970 report on documentation policies at SAIL[394]. This document outlines laboratory genres as a multi-stage system, from the most formal, stable, and outward-facing genres to the most non-formal, fluid, inward-looking ones.

To start with, SAIL members communicated their inventions and scientific results to the world in quite traditional scientific ways: through publications and conference presentations. Further, within the laboratory, project outputs and findings were communicated in the genre of AI Memo (AIM).[395] AI Memo was a rather formal genre, organized according to the standards of scientific publications, accompanied with references, bibliographies, etc. Typically, AI memo presented significant research results: for example, doctoral thesis or findings of a research project might have been published as a Memo. Memos were to be printed, approved and signed by research supervisors, and cited in reports to funding agencies.

In 1966 the genre of SAILON – SAIL Operating Notes – was introduced [396]. Compared to Memos, it was a more informal genre, describing technical systems, programming languages, and software rather than academic findings. Targeted for use by various projects both within and outside the lab, SAILONs constituted manuals and user guides on the laboratory's major technological developments. Like Memos, they were printed and signed by executives, but in contrast to Memos, they occasionally got outdated and require revisions.

394 https://www.saildart.org/DOC.LES[S,DOC]

395 The Memo lists for different years are stored in the directory [BIB, DOC]. The most recent one is for 1984 [https://www.saildart.org/AIMLST[BIB,DOC].

396 Some SAILON files can be found in the directory https://www.saildart.org/[BIB,DOC]/.

Lastly, around the early 1970s, yet another genre appeared: programming notes, which were "used to describe small system programs or user programs that have not yet been released for general use"[397]. Unlike Memos or SAILONs, through which the laboratory presented inventions, breakthroughs, or new technical systems to itself and to the world, programming notes were oriented exclusively towards the lab members. They outlined pieces of new software, small utilities, bits of code, or possible workarounds for existing programs to be used by the SAIL programmers. Program notes form a common and unified technical framework of the lab, which allowed one to use the existing code or software instead of searching for new solutions and developing new code for each new task.

As Les Earnest summarized in describing SAIL's "three-tiered" scientific genre system,

```
our A.I. Memos are nearly all large and stable. SAILONs are both large and
small and mostly unstable. Program Notes are mostly small and unstable[398].
```

Being "unstable" and constantly updating, programming notes were usually not printed, instead being stored as digital files in the system. So unlike AI Memos or SAILONs, only single printed copies of which survived in the Stanford archive, digital and unstable programming notes were perfectly and thoroughly preserved in the SAILDART.

To keep them, a "fictional" programmer, DOC, had been created, aggregating a number of directories – public repositories of programing notes created by and for everyone. This communal, generic fictitious programmer DOC[399] was a kind of alter ego of the lab's community, bringing together files of different programmers from different directories, in a single space accessible to all. DOC's directory stored about three thousand of programming notes covering compilers and preprocessors, programming languages, Mail and News services, small utilities.

Programming notes were the most informal of all the genres listed above; they were neither controlled nor authorized by anyone, nor did they follow any formal rules. The notes often even had no headings or authors specified. The authorship of a file could be deduced only by its extension, which repeated the username of the programmer. A programming note, thus, did not present the author's innovation, but served as a form of contribution to the common cause, to the shared arsenal of technical tools accumulated in the laboratory. Any lab programmer who developed a program that might have been useful to someone else in the lab,

397 https://www.saildart.org/DOC.LES[S,DOC]
398 Ibidem.
399 https://www.saildart.org/DOC

described it and published it in the DOC shared and collaborative space. As the program changed, the programmer (not necessarily the original author of the program) published its update. The DOC directory thus formed a "knowledge commons" (Hess, Ostrom 2011), a free, open, collaborative and communal space shared and fleshed out by the members of the lab.

Rather than simply informing the community about some new utility, DOC was aimed at integrating it into practices. What it stored was instrumental knowledge intended for use, application in the course of scientific everyday life. Programming notes were constantly updated, and these updates were tracked due to the automatic notifications system. Thus, DOC was not a formal structure prescribing certain procedures to be followed. To use de Certeau's terms (de Certeau 2011 [1980]), it stored not *strategies* imposed from above, but *tactics* proposed and practiced by the community members themselves. The archive, then, affords the opportunity to glimpse not only communal discourses, but also these very techno-collective sharing practices.

4.5.4. PUB: the community engages in practice

Beyond the shared and communal "instrumental knowledge", one can uncover even deeper structures of collectivity in the archive: those embedded in the actual SAIL practices. I will discuss them through the example of one forgotten SAIL innovation – PUB.

PUB standed for PUBlication language; it was a "document compiler"[400] designed to turn "manuscripts" (text as presented in a text editor) into "documents" (text formatted for printing or display on a terminal). PUB provided the possibility to number pages, mark headings and footings, create table of contents, generate cross-references and many other features that today are relegated to Microsoft Word, PHP and HTML/CSS.

PUB was developed by Larry Tesler, a programmer best known for creating the cut/copy/paste command. In the late 1960s, Tesler worked on PARRY, but got disillusioned with Artificial Intelligence, and sought to focus on user interface design[401]. Les Ernest then offered him to undertake the document compiler project. The first version of the program was released in late 1971, followed by several more versions and modifications. In 1974, Tesler left SAIL for Xerox PARC, but PUB continued to be used at the laboratory through the 1980s until the famed TEX, developed by Donald Knuth, took over.

---

400 Today we would rather call it mark-up language.
401 For Tesler's biography and recollections cf. Tesler 2012, Markoff 2006, 130-134.

According to John Markoff, PUB "was a great success" and "foreshadowed HTML" (Markoff 2006, 133), yet today it appears to be an all but forgotten page in the history of digital publishing and markup. Notably, for example, Matthew Kirschenbaum, who devoted an entire book to the history of word-processing (Kirschenbaum 2016) and made Tesler one of its main characters, never once mentions PUB. Tesler himself, however, returned to the document compiler and even provided his PUB SAILON with a detailed (albeit unfinished) commentary comparing PUB functions with other languages and technologies[402].

Debugging as a dance of collective agency

In the early days of interactive computing[403], the debugging process was the most time-consuming part of the work involved in writing a program; it took many weeks and was a constant subject of refinements and improvements. Donald Knuth, for example, was debugging his TEX on nights when the computer was relatively free, and kept a detailed journal of the process[404]. According to his diary, he spent most of the nights of March 1978 debugging TEX, working, on average, 7 hours each night[405] – it took that long just for the initial debugging phase.

Tesler wrote the first version of PUB in 1971, but the process of debugging and refining the program continued for years. In 1975 Les Earnest published in the DOC directory the file PUB.UPD, "a collection of facts, folklore, and fables" about the program[406]. The document began as follows:

```
While Pub may be a very large and slow program, it compensates by being
idiosyncratic. If you think it is buggy now, you should have seen it
earlier[407].
```

The very process of debugging is a perfect example of what Andrew Pickering called "the dance of agency" – the interaction between human and nonhuman, in which they in turn take on an active or passive role (Pickering 1995). Pickering's prime example is his observations of how Donald Glaser developed the bubble chamber. Glaser's interaction with the machine resembled a dance in which Glaser first watched the machine in action, then made

402 https://www.saildart.org/allow/PUB/pub_manual.html

403 It was interactive computing that made possible face-to-face debugging, in which a programmer can observe the behavior of a program without intermediaries. Cf. McCarthy et al. 1963.

404 https://www.saildart.org/DEBUG.LOG[TEX,DEK]4.

405 Knuth then turned this experience of observing the debugging process into an article: Knuth 1989.

406 This file was updated until 1985, when the PUB was finally replaced by TEX and Earnest raised the issue of its preservation.

407 https://www.saildart.org/PUB.UPD[S,DOC]1

some adjustments to it, then observed it again and so forth until a working bubble chamber came into being. The practice of science thus appears as an interplay of (human and nonhuman) agencies in which they mutually adjust to one another.

The debugging process follows exactly the same scenario: the programmer observes the action of the program, when it gives a bug, they modify some piece of code and run the program again up to the next problem, and the whole dance is performed once again. In the case of PUB (and many other SAIL programs), the specificity of the "dance" lied in its collective nature: debugging PUB took place not between scientist and machine, but between scientific collective and machine.

SAILDART makes it possible to witness this collective dance. The bugs were discovered and reported by various members of the lab. Tesler had more than three dozen bug messages[408] in his directory from various users of the system who had come across bugs in the process of running PUB:

```
Dear Larry,
       Got another PUB bug for you. I tried OMAN.PUB with indexing turned
on. I used 12K string space and it appears to work! (that's the good news.
Now for the bad) When it began PUB PASS TWO, I got the error message
INVALID index No. 1 to ARRAY LABTAB. However, I had a parity error in my
core image during the 4 hours it was (time shared) in core. I finished the
pass one files, and they appear okay. CPU time until the bug: 1:08:53 (a
new record!!).
                              -Kurt[409]
```

```
THE FOLLOWING LINE IN MY FILE SEEMS TO CAUSE PUB TO LOOP FOREVER:
.TITLE AREA HEADING LINES 1 TO 3 CHARS 6 TO 65;
THE FILE ALSO CONTAINS THE LINE:
.EVERY HEADING(GLO.THE,STRATEGY OVERVIEW,{DATE});
                                               09-JAN-73 1405        1,GG[410]
```

```
08-JAN-73 1053         2,KKP
PUB bug: when I say "R PUB" then "ARPA.PUB[P,LES]/D", it runs to page 16,
then gives "ILL MEM REF AT USER 402507". DCS looked at it and said he
thought you were concatenating off the end of core. I'm thoroughly hung
up[411].
```

In this way, Tesler as a developer received feedback from *both* the computer and other lab members, who also became involved in the dance of agency. They observed the operation of the machine, detected a bug, and then communicated it to Tesler, who implemented changes and then the *collective* observed the way the program performed again up until the next bug.

408 Files with .MSG and .BUG extensions in Tesler's directory https://www.saildart.org/TES..
409 https://www.saildart.org/FEB.MSG[2,TES]
410 https://www.saildart.org/GG.MSG[2,TES]
411 https://www.saildart.org/BIG.MSG[2,TES]

Yet the modes of community participation were not limited to finding bugs. Tesler was also receiving input from his colleagues on how to improve or reconfigure the compiler. Notably, the lab members did not simply point out the lack of certain features, but offered the very pieces of code to be added or modified in the program:

```
27-MAR-73 1112          LDE,DCS
Larry,
        Consider the following PUB macro:
.MACRO PAR (TEXT)
.⊂ SOME CONTROL STATEMENTS;
↓_TEXT_↓
.CONTINUE ETC. ⊃
        If I want underlines to go all the way across underneath TEXT, I
have to put "_" chars in for " " throughout TEXT. This means changing
all calling sequences if I change my mind about the underlining later.
If you agree, praps we should put a feature like this in the list of
good PUB things to do.
                                                                 Dan[412]
```

Some of these messages suggest modifying the PUB so that it would have better fit into the existing ecosystem of the laboratory. For example, in the message below, REG (Ralph Gorin) explained to Tesler how to connect the existing text-to-print (SPOOLER) mechanisms to the compiler:

```
∂24-OCT-73 1555         ACT,REG
I SUGGEST THAT YOU IMPLEMENT THE FOLLOWING MODIFICATIONS TO PUB.
1. (THIS ISN'T ANY WORK FOR YOU) LOAD PUB (AND PUB2) WITHOUT THE SAIL
SEGMENT. THIS WILL INCREASE THE EFFICIENCY OF RUNNING PUB.
 2. WRITE THE NAMES OF THE FONTS (AND SPECIAL XGP SPOOLER SWITCHES)INTO THE
FIRST PAGE OF THE DOC FILE. THE XGP SPOOLER WILL READ FONT NAMES FROM THE
FIRST PAGE IF YOU USE THE COMMAND: XSPOOL <FILE>/XGP (BETTER THAN THAT, USE
THE EXTENSION ".XGP" FOR THE DOCUMENT FILE. THIS EXTENSION SIGNIFIES TO
XSPOOL THAT THE FONT NAMES WILL APPEAR IN THE FILE.)(CAUTION: ONLY FONT
NAMES AND XGP SWITCHES (LMAR,PMAR,BMAR,TMAR,RMAR AND XLINE)ARE LEGAL IN THE
ENTIRE FIRST PAGE.[413]
```

Lab members were involved not only in finding bugs, but also in correcting and adjusting the source code of the program. PUB thus appeared as a product of collective work and *collective* interactions with the machine. With Tesler departed, the compiler became the responsibility of the entire lab and no one in particular. Further debugging of the program was distributed among community members and its fruits – a variety of collectively found solutions and tricks on how to interact with the program – were reflected in a number of documents in the DOC directory[414].

[412] https://www.saildart.org/MAR.MSG[2,TES]
[413] https://www.saildart.org/NOV5.MSG[PUB,TES]
[414] https://www.saildart.org/PUB.TES[S,DOC]1

Debugging PUB thus appeared as both a dance with the machine and as a collective, social dance, realized through both social and technical operations. In this complex dance, the machine and different members of the lab took turns carrying agency and were mutually adjusted to one another. Observing the debugging process via the archive, one can trace how SAIL's collective practices are woven into the very technologies of the lab.

Collective practices of PUBbing

Our writing tools are working on our thoughts.
Friedrich Nietzsche[415]

Created in a negotiation between a collective and a machine, PUB was then put into use by that very collective and came to define its word-processing practices. The compiler almost instantly became an integral part of the practices and economies of writing, first at SAIL and eventually beyond the laboratory[416]. It was used for formatting laboratory Memos, SAILONs and reports, for publications, and for a vast number of doctoral dissertations.

As we have seen, PUB was the fruit of collective work of the lab. Moreover, like many other pieces of software, it was originally designed for SAIL systems, based on the encodings, keyboards, and printers of the laboratory: for example, SAIL characters "alpha" and "beta", incompatible with common ASCII encoding, were used as PUB commands. As such, PUB was an outgrowth of the social and technical systems of the laboratory.

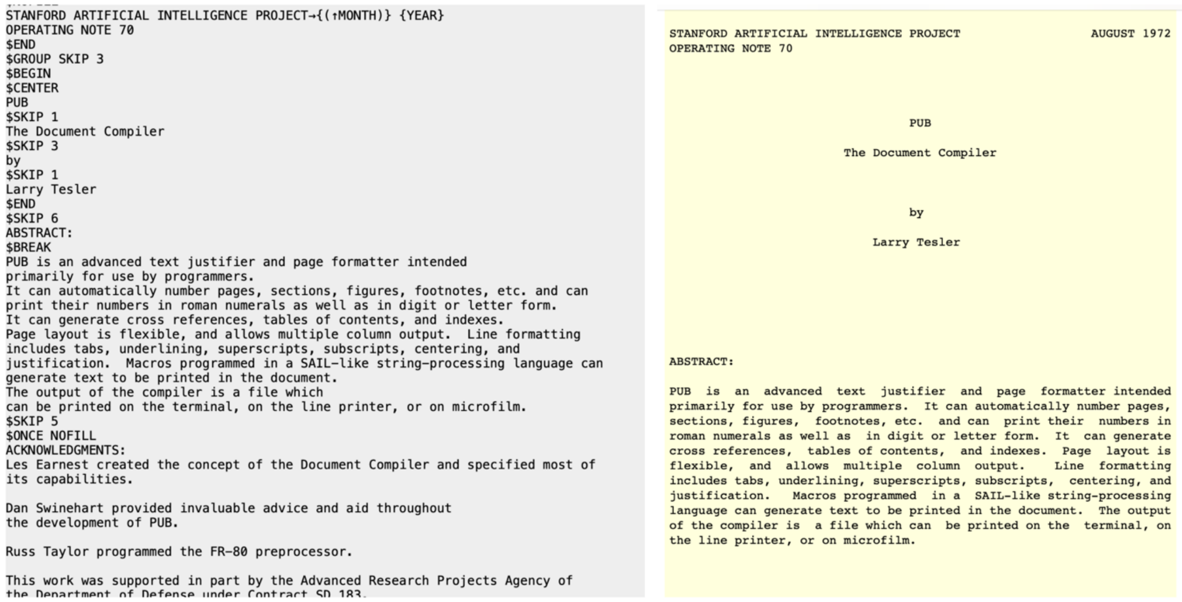
For another thing, like any software, PUB in turn configured its users and their practices (Woolgar 1990). The compiler prescribed certain rules for text formatting, starting from the width of the page (69 characters) to the special characters that were to be used and those that were considered "illegal"[417]. It also defined a common view of the document structure and the markup: such as that the document structure was hierarchical, with "portion" being the main unit of reference. But perhaps most importantly, PUB prescribed a certain technology of document production. Like a mark-up language, it merged the text to be formatted and the "formatting" code in one file ("manuscript") (Fig. 4.17). This way, the code and the writing within the document intermingled, flowed into one another. Further, the compiler implied a two-phase system for turning a "manuscript" into a "document". From the original .PUB file

415 Cited in Kittler 1999, 200.

416 In the archive one can find Tesler's negotiations on modifying the PUB for other printers and systems.

417 A discussion of these and other technical details can be found in Tesler's PUB manual: https://www.saildart.org/allow/PUB/pub_manual.html

containing the code, the .PUG, .PUZ and .PUI files were first generated, and then the "final" document – a .DOC file – was made. In the lab, this process of converting from .PUB files (text with code embedded in it) to a .DOC file (the finished, clean version in which the markup was not visible) was called PUBbing. All changes to text or code were made at the manuscript stage, and PUBbing was only performed when the text was completely finalized. The programmer thus could not see the final version when working with the document, but only imagined it until the conversion was done – much like when writing a program script.



```
STANFORD ARTIFICIAL INTELLIGENCE PROJECT→{(↑MONTH)} {YEAR}
OPERATING NOTE 70
$END
$GROUP SKIP 3
$BEGIN
$CENTER
PUB
$SKIP 1
The Document Compiler
$SKIP 3
by
$SKIP 1
Larry Tesler
$END
$SKIP 6
ABSTRACT:
$BREAK
PUB is an advanced text justifier and page formatter intended
primarily for use by programmers.
It can automatically number pages, sections, figures, footnotes, etc. and can
print their numbers in roman numerals as well as in digit or letter form.
It can generate cross references, tables of contents, and indexes.
Page layout is flexible, and allows multiple column output.  Line formatting
includes tabs, underlining, superscripts, subscripts, centering, and
justification.  Macros programmed in a SAIL-like string-processing language can
generate text to be printed in the document.
The output of the compiler is a file which
can be printed on the terminal, on the line printer, or on microfilm.
$SKIP 5
$ONCE NOFILL
ACKNOWLEDGMENTS:
Les Earnest created the concept of the Document Compiler and specified most of
its capabilities.

Dan Swinehart provided invaluable advice and aid throughout
the development of PUB.

Russ Taylor programmed the FR-80 preprocessor.

This work was supported in part by the Advanced Research Projects Agency of
the Department of Defense under Contract SD 183.
```

STANFORD ARTIFICIAL INTELLIGENCE PROJECT AUGUST 1972
OPERATING NOTE 70

PUB

The Document Compiler

by

Larry Tesler

ABSTRACT:

PUB is an advanced text justifier and page formatter intended primarily for use by programmers. It can automatically number pages, sections, figures, footnotes, etc. and can print their numbers in roman numerals as well as in digit or letter form. It can generate cross references, tables of contents, and indexes. Page layout is flexible, and allows multiple column output. Line formatting includes tabs, underlining, superscripts, subscripts, centering, and justification. Macros programmed in a SAIL-like string-processing language can generate text to be printed in the document. The output of the compiler is a file which can be printed on the terminal, on the line printer, or on microfilm.

Figure 4.17. The document on the left is PUB.PUB [https://www.saildart.org/PUB.PUB[2,TES]1], Tesler's SAILON manuscript on the document compiler. One can see here how the text and the code coexist in the same file. On the right is the printed final version of the same page [https://www.saildart.org/allow/PUB/pub_manual.html].

PUB thus brought the process of document processing as close as possible to that of programming. For the first time, the programmers and not the secretaries were in charge of formatting their own writings. PUB (as well as other, not so "idiosyncratic" compilers at SAIL) thus changed the very outlines of the programming profession. Not for nothing, therefore, was it conceived as an indispensable part of the office of the future envisioned by John McCarthy[418] and Les Earnest. Curiously, in the software discussion we started with, it is the document production practices that were recognized as successful by all the programmers, while other languages and systems were disputed. Designed by the community, PUB also established the shared rules, concepts, and practices that the community follows.

From joint discussions to collaborative practices, DOC, PUB, SOFT each capture and articulate the SAIL collective identity in its own way. Taken together, however, they manifest

418 https://www.saildart.org/OFFICE.2[W80,JMC]

SAIL's communality in a very different manner than the way it is constructed in community stories or in the social sciences reflection. What they bring to light is how the SAIL community is produced through technical and social practices, the remnants of which can only be found in a "perfect" machine archive. Through the SAILDART, we can thus delineate the SAIL community as a "recursive public" (Kelty 2005) that defines and makes itself through the shared material practices.

Conclusion

The example of the reaction key archive discussed in the previous part and that of SAILDART offer two very different, almost opposite configurations. While Part III meticulously models the archive, diligently establishing connections between the archival fragments, Part IV deals with an entirely different type of archive: not orderly and structured, but chaotic and cluttered. If the method in the previous part is to weave archiving fragments into a "collage", this part is more about sifting through an abundance of existing remnants. It confronts complex, technologically dense, and multiply mediated objects that necessitate additional layers of technological interpretation.

These two scenarios – the digitized and meticulously modeled archive on one hand, and the "natural", almost fortuitous, born-digital archive on the other – illustrate the vast diversity inherent in digital archives. They reveal how differently these archives can manifest, the varied themes they can unravel, and the need for distinct methods, engagement techniques, and interpretive approaches.

Through the example of SAILDART, one can see how the ideas of technological memory, the total and "ideal" archive, that were once confined solely to the realm of historians' imaginations, are coming into play. These configurations (not so "ideal" in practice) obviously call for (historical) reflection and pose new challenges for both historians and archivists.

As we have seen, born-digital archive implies distinctive mechanisms of memory and knowledge transmission: those ways of selecting, transmitting, and preserving historical artifacts that have been in place since the nineteenth century no longer work or work somehow differently. The very notion of memory (and oblivion) that underpins the traditional archive is being reshaped and altered, becoming inextricably intertwined with the technological *dispositif*. In sequentially examining the media archaeology, mediology, and exegesis of SAILDART, we consistently encountered increasingly complex layers of technological mediation. The deep intervention of technology in the processes of memory and transmission – arguably the main hallmark of the born-digital archive – changes both how we encounter the past through an archive, and the way we make sense of it. Technological memory, as we have seen, puts the question of understanding back on the table and demands ever new (technological or machine-based) forms of interpretation. In the scenario with the reaction key, modelling is merely one potential avenue of interpretation. Contrastingly, SAILDART vividly exemplifies the notion that at times, our capacity to understand, preserve, transmit, and even create is entirely contingent upon technological means.

We have also seen the frontiers of what the born-digital archive can testify to. At the end of the day, it offers insights into the digital practices of the past: digital writing, reading, communication, programming... Further, through the remnants of practices, it leads us to broader vistas: questions about how the private and the public, the individual and the collective, the everyday and the professional are shaped through the digital. As we have observed through comparisons of the archive with other commemorative practices, what it primarily preserves is the interaction of the user with technology. This interaction, in turn (particularly for the culture of early AI), becomes a key to understanding how users communicated with each other. It sheds light on specific types of digital communal practices and the "technological imagination" of the community.

The analysis of SAILDART performed in this part addresses, in a sense, both the past and the future. It looks to the past as it delves deep into one of the earliest born digital archives, filled with remnants of digital practices and technological imaginings. It faces the future, as many of the conclusions drawn from the SAILDART example will clearly be relevant to the new (probably immense) digital repositories we will be dealing with in the coming future.

In this sense, dealing with the unintentional and involuntary residues of SAILDART, this part also sharpens the question of how scientific institutions can/should anticipate their future archive and shape their future collective memory in the present. As historical as "modern" consciousness is, the present-day routine is hardly recognized as valuable. This is perhaps particularly characteristic of science oriented towards the future, and even more so of Artificial Intelligence, dealing with the supposedly eternal.

Conclusion

> When we endeavor to examine the mirror in itself, we discover in the end that we can detect nothing there but the things which it reflects. If we wish to grasp the things reflected, we touch nothing in the end but the mirror. — This is the general history of knowledge.
>
> Friedrich Nietzsche

From the 1980s onward, Science studies have shifted their focus towards "science-in-the-making", unsettled and full of uncertainties, as opposed to the "ready-made science", mature, stabilized, and taken for granted. It is in this state of uncertainty, the "in-the-making" phase that the formation of scientific knowledge becomes visible, allowing the researchers to observe how the established facts of "ready-made" science emerge.

The *past of science* is not given to us as "ready-made" either. It is not a static, unchanging reality waiting to be uncovered. Rather, as the constructivist historiography argues, it is being continuously made, re-constructed and re-presented. The way we perceive the past of science, as Stuart Hall remarked on a different matter, is "an ongoing project, under constant reconstruction. We come to know its meaning partly through the objects and artefacts which have been made to stand for and symbolise its essential values" (Hall 1999, 5).

This dissertation tackles this "past-of-science-in-the-making", delving into how knowledge, representations, and imaginations of science's past are being shaped. It does so by zooming in on one specific locus where the past of science "is being made": the digital archive of scientific residues.

Such an archive brings to the fore certain objects in such a way that they represent and "stand for" the scientific past. The selection of these objects, the way they are represented, what they symbolize and how they are interpreted are far from being natural or "ready-made" processes. As we have observed, these processes are entangled with various regimes of value and different politics of memory, involving an interplay between the innovative and the obsolete, between scientific usage and patrimonial value. In this continuous project, the archive does more than store some objects of the past; it becomes an active agent in their interpretation.

This dissertation views the digital archive of science as a phenomenon situated within, and emerging from, the simultaneous interplay of certain historical and technological horizons. On the one hand, it appears as an outgrowth of the of the current "regime of historicity" (Hartog 2003), the memorial/patrimonial turn with its sensitivity to the heritage of the past and along with the growing interest in the everyday life of science. The archive of residues comes into play when Science studies become interested in the tangible everyday scientific artifacts, when these objects are found to possess heuristic value and come to "symbolize and stand for" science-in-the-making.

On the other hand, the digital archive (of science) functions as a particular kind of technical configuration, or more precisely, an ensemble, emerging from the integration of numerous technologies and information systems. The way archival data is represented, interconnected, and presented to the user is shaped and mediated by this ensemble. As a distinct, multilayered, and complex technological infrastructure, the digital archive possesses its own affordances and constraints and embodies certain assumptions about how information is stored and represented.

As both historical and technological *dispositif*, the digital archive of science offers particular forms of "making the past". This dissertation probes into how we understand and interpret the past through such an archive, focusing on its peculiar modes of representation, the methods of treating the past it offers, and its transmission mechanisms that bridge the past and the future.

Archive and representation

The digital archive presupposes a particular regime of representation, through which the user encounters objects of the past. As we have seen through the distant reading of 118 collections, this order is radically different from that of the traditional physical archive and its classical principles of storage. The digital archive shifts away from emphasizing the authenticity, materiality and "presence effect" of the past. Instead, users engage with digital representations of scientific artifacts, which are arranged in a certain fashion.

Rather than centering on the individuality and uniqueness of each item, the digital archive emphasizes the abundance, the multitude, the *masses* of things brought together across *fonds* and physical repositories, disciplines and narratives. These emerging economies of scale are made possible not just because digital technologies facilitate the aggregation of artifacts and formats beyond physical confines, but also due to the intrinsic properties of the digital object. Composed of bits and bytes, it can be broken down into discrete, individual components, each in its turn autonomous, mobile, manipulable, and capable of forming diverse combinations. Therefore, the digital archive is in a sense fractal: it no longer preserves the integrity of funds and the wholeness of things, but deconstructs them into parts, thereby storing a potentially infinite array of fragments. Within this digital mosaic, every single fragment can attain a prominence on par with the complete artifact, redefining the very concept of archival value.

A new object of historical representation and imagination thus enters the scene: a *multitude* of (scientific) objects. Visually, this *multitude* is manifested in what I have called the ornamentality of the archive: the objects together form a certain rhythmic pattern in which the unique becomes indistinguishable and the quantitative aspects, the rhythm of the pattern, come to the fore. The *multitude* offers new forms of encounter with the past and new methods of studying it: distant reading as opposed to close-reading; macroscopic thinking as opposed to object- or *fonds*-scale thinking.

As such, the digital order challenges the traditional forms of historical representation and contemplation: linear, sequential and perspectival. Instead, it enables a mode of historical engagement characterized by non-linearity, simultaneity, absence of depth, and the amalgamation of various temporal sequences and historical logics. It also alters the forms of interaction between the historian and the past: shifting from an intimate and tactile experience of engagement with unique sources to various forms of "virtual manipulation" with the ornaments of things. Similarly, for the public, the past becomes visible through, and mediated by, the digital interfaces, that are already familiar from other digital practices (as we have observed, the interfaces of digital collections often bear a resemblance to those of many existing web platforms, such as Amazon).

The digital archive enables the manipulation and operation of this *multitude* as a cohesive entity. In contrast to the traditional archive, which maintains the "original" order, it introduces a plurality of orders. Here, masses of objects can be (re)situated, (re)combined and (re)arranged in various combinations and configurations. Each configuration orchestrates the *multitude* of archival objects in its own way, reflecting diverse principles of what these individual items have in common.

The ways in which these objects are combined depend on the indexing systems employed within the archive and the methods it uses to establish connections between items. The indexing systems and connections in place not only determine what users can find in response to their information needs, but also how the objects will be interpreted, understood, and ascribed meaning within the context of the collection. They determine what is visible and what is hidden, build implicit hierarchies, set connections and fractures.

As we have observed, even an object as commonplace as a light bulb can take on a multitude of meanings, depending on how it is presented and contextualized within the archive.

The process of semantization appears to be especially sensitive where special knowledge is needed to interpret an object, such as in the case of the residues of science. Preserving such artifacts implies not only conserving their material shell, but also framing the context of their meanings, uses, and interactions with other objects, as well as individuals, institutions, disciplines, and concepts.

The digital archive introduces a paradigm where there is no single, definitive order, but rather where multiple orders are possible. Thereby the very concept of order inherent in the archive is being reconfigured: no longer one finds the only possible order, whether natural or organic; instead, the order becomes malleable, mutable, and open to *modelling*.

Archive and modelling

When modelled, the digital archive takes shape depending on the intentions, interpretative lens, or research question of the historian/archivist/digital curator. Historians as part of their research can themselves conceptualize the archival structure, annotate its contents, and map out the web of relationships and connections between items in such a way that the archive can produce new knowledge.

By engaging in data modelling, historians craft a specific representation and interpretation of historical phenomena, woven into the fabric of technical systems and algorithms. Such an archive no longer operates on its own inherent logic. It transforms from a repository where historians get sources to be referenced in a historical narrative, into a form of historical representation and articulation in its own right. As such, the archive gains a creative (generative) power, the capacity to bring forth new meanings, interpretations and new knowledge.

To illustrate the operation of archive modelling and its epistemic capabilities, I built a prototype semantic model for the archive of scientific residues. The process of construction, embodying the synergy of "thinking-in-doing" and "doing-in-thinking", revealed both the diverse challenges inherent in this method and the interpretative possibilities it opens up.

In terms of epistemology, no matter how fruitful ontologies might be, they also bring the danger of universalization and naturalization of knowledge. As such, constructing a semantic model requires an ontological commitment acknowledging the model as merely one among many possible representations of the past, neither natural, nor definitive. Every ontology, including the one presented here, comes with its constraints, limitations, and biases.

Therefore, the boundaries and rationale behind each model need to be clearly articulated and transparent.

Furthermore, one has to realize that any model is bound by its time, reflecting only the *current* state of knowledge. Accordingly, the ontology I propose recognizes that the contextualization of scientific remnants within a digital archive should be undertaken from the perspective of current developments in various fields of knowledge, including Science, Material and Cultural studies. In other words, I propose to weave contemporary *theory* into the archive assuming that the interpretation of scientific residue is only possible based on current interpretative perspectives.

To put the ontology in action, I have "modelled" an archive centered around a minor device from experimental psychology – the reaction key. The outcome of this experiment is a series of sketches towards an object-oriented history of the reaction key. This history, crafted from the perspective of this particular minor instrument, interweaves artifacts, their archival representations, people, things and institutions, concepts, technical and epistemic objects, and even different scientific paradigms. Such a history is very different from the traditional linear historical narrative: it branches out in various directions, forming diverse, non-linear, and rhizomatic connections. It addresses a diverse range of questions about the instrument's history: from the types of instruments alongside which it was used within various experimental setups, to the specific movements it was designed to capture.

Opening the objects of science for distant reading, this history renders visible some patterns, tendencies or continuities that have been invisible without it (e.g., the genealogy of the "technological tradition" of the reaction key). It can also illuminate, elaborate on or challenge established linear narratives within the history of psychology or regarding specific instruments.

As a form of historical work, modelling is quite different from the "traditional" methods of conducting historical archival research. It almost literally enacts the idea of *making* history. It entails classifying, arranging and linking together disparate historical fragments, residues, and snippets scattered across various archives. Modelling in historical research thus appears as a form of bricolage, a creative process where the historian intricately clips, carves out, and assembles details from the past. In the act of classifying these pieces, a new historical collage is created. During this activity, classification becomes a form of conceptualization; the acts of classifying (*classer*) and thinking (*penser*) are intertwined into one unified process, *Penser* /*Classer* (Perec 2003 [1985]).

Archive and transmission

The archive serves not only as a means of representing and contextualizing objects from the past for the present but also raises questions about their transmission into the future. This issue is particularly pertinent for born-digital collections, which lack physical, fixed, and stable counterparts. In this case, again, the digital introduces some notable alterations to the 'traditional' mechanisms of archival transmission.

I have explored this new logic of transmission through the case study of one of the earliest born-digital archives of scientific residues – the archive of the Stanford Artificial Intelligence Laboratory. As such, SAILDART offers a rare glimpse into how born-digital archives can be understood and interpreted half a century after their inception.

As we have observed, in the born-digital archive, memory and technology become inextricably intertwined and inseparable from each other. This convergence leads to a reconfiguration of the very concepts of memory (and oblivion) that constitute the core of the traditional archive.

In the case of SAILDART, we encounter an archive that is compiled, catalogued, annotated, stored and maintained by machines. More than that, the very preservation and recall of born-digital residues, decades later, are made possible by numerous operations of conversion into different media forms and numerous acts of *réécriture*, as opposed to maintaining an unchanged (authentic) state of record in a traditional archive. SAILDART, thus, presents a radically new configuration where the computer emerges not just as a vessel for memory but also as its mediator and even the agent of memory work.

The result is this new kind of archive, either an "ideal chronicler" capturing every minutiae with unerring accuracy, or a chaotic dump of undifferentiated noise and debris. As exemplified by SAILDART, such an archive operates without historical or archival selection, making no distinction between what is significant or trivial, valuable or not. As such, it rekindles the utopian idea of a total archive that records everything "exactly as it was" and envisions an "ideal", objective archivist, which turns out to be a machine.

In discussing the epistemological potential of such an archive, we observed how it brings our gaze to the very everyday noise of the early AI and the material remains of the technology of the time. But more than that, it also allows a view into how these technologies were utilized and how users interacted with them (and each other) in the past. The archive appears as a form of both machine and collective memory offering a glimpse into the experience of early interactive computing, the way the community communicates, shapes and defines itself through technological practices. It provides insights into, for example, the (digital) scriptural economy of John McCarthy, as well as his intricate system of communication as it was structured and used at the dawn of this technology. Or it allows us to observe the early collective debugging practices, the collaborative interplay between the community and the machine.

It enables us to see not only how technologies were used, created, and written but also how they were imagined at a specific point in time. The archive provides a lens to view old technologies at a time when they were at the forefront of innovation. It allows us to explore the then-new technologies and to glimpse the collective imaginings of the future as they were perceived at that time. We can thus trace the horizon of technological imagination of the time, seeing not just the precursors of today's technologies, but a tapestry of potential futures, an "*histoire des possibles*".

Archive in the making

If the angel of history looks backward into the past (Benjamin [1940]), the angel of the archive casts its gaze in both directions simultaneously: towards the past, which it preserves, and into the future, for which it is preserving.

In between the past and the future, the archive, however, is deeply shaped by the present. The selection between valuable and non-valuable is made through the lens of current notions of value; the way the past is understood and addressed for the future is rooted in the current (technological, historical, epistemological) horizon. Therefore, it sharpens the question of how we understand objects from the past within the context of the present and how to re-frame this understanding for the future.

The digital archive meets this concern by reimagining the very concept of preservation: it is no longer about safeguarding an object in its unaltered, fixed state or maintaining an immutable order of things as they were in the past. Instead, preservation becomes an ongoing process of reconfiguration, reactualization, and reconstruction of the archive to align with the user's horizon of understanding. The digital archive, in other words, is permanently "in the making", dynamically interacting with the present, and redefining our relationship with the past.

Its plasticity and dynamism are particularly evident at the technological level. As the materiality and stability of the object are no longer present, storage turns into an endless succession of *réécritures* – a series of media conversions and mediations designed to render "information" comprehensible and accessible. However, it is equally important to recognize that the new paradigm of a "modelable" archive is also susceptible to conceptual and value-based changes dependent on prevailing research paradigms, memory politics, and value regimes.

In the digital archive, the residues of past innovations thus again become subjects for re-invention and re-contextualization. The very notion of what constitutes a residue, and its perceived value, is in a state of flux, influenced by changing uses, perceptions, and interpretations. Furthermore, as we have observed, understanding technologies of the past sometimes necessitates the application of "new technologies". For instance, this is the case with old fragments of code from SAILDART, which can only be comprehended through additional layers of technological mediation (such as ChatGPT). This archival dynamic adds a layer to the dynamic of innovation and obsolescence, the "creative destruction" that opened this dissertation.

The malleability and reconfigurability of the archive offer a wide array of opportunities for engaging with the making of the past. It opens doors to experimenting with various forms of interpretation, semantic enrichment, and narrative construction. In this context, the digital archive can serve as an excellent educational resource. It would allow students to actively participate in the "past-of-science-in-the-making" by dynamically modelling, re-ordering and contextualizing the archive. Such an exercise would not only provide them with practical experience in handling historical data but also foster critical thinking, interpretive skills, and a deeper understanding of how historical narratives are shaped. The flexibility of the digital archive also present a valuable opportunity for scientific institutions to engage their communities more actively in memorialization efforts, for example, enabling them to directly model, annotate, and shape their archives. And of course, as has been repeatedly discussed in this work, the digital archive appears as a valuable addition to a historian's toolkit enabling historians to craft multiple histories across and beyond physical repositories.

That said, the very idea that the archive is fabricated, crafted, and expresses a particular (theoretical and therefore ideological) perspective may seem disturbing or even subversive. What can history rely on if its main pillar, the archive, is constantly in flux and can no longer provide the sought-after permanence? In the age of the polyphony of social media, fake news, the ever-present conflict of narratives and interpretations, it might seem that the archive needs rather a defense of its authority than the constructivist view that this study offers.

Highlighting the constructed nature of the archive and advocating for its modelling, this research does not imply a descent into relativism, nor does it suggest that the past is entirely pliable and subject to arbitrary modification. On the contrary, it calls for a critical awareness of the epistemological foundations of the archive, for a clear understanding of how the archives are being modeled and where precisely the boundaries of the representation they offer lie.

Insofar as the digital archive presents specific modes and orders of representing the past, and given that the engagement with the past, for both historians and the public, is becoming increasingly mediated by technology, it is imperative to understand the intricacies of these numerous mediations that have become essential for understanding and remembering. In other words, the "archival fever" (Derrida 1998) needs to be cured, the "effect of reality" (Farge 1997 [1989]) needs to be dispelled, the invisible infrastructure needs to be made visible.

Expanding the scope of inquiry, the same conceptual approach – examining how technology mediates understanding – can be extended beyond the realm of digital archives (of scientific residues). It seems equally applicable to a variety of infrastructures that are often overlooked and remain invisible: from commercial platforms like the aforementioned Amazon which clearly propose specific relationships with things, to the websites of universities and laboratories that convey certain notions of knowledge.

That said, exploring how the past is made through the digital archive is also not limited to the aspects and methods explored in this dissertation. While this study focused on the role of objects as "residues" and "witnesses", exploring representations of oral history within the digital archive presents another avenue for exploration. This could involve examining how personal recollections are captured, shaped, and mediated through digital technologies. One more aspect of interest, only briefly touched upon in this study, is the question of emulation and re-enactment in relation to understanding and technology. Along with modelling, emulation appears as one of the main new techniques of preservation and understanding of the past. A comparison of the two methods, their limitations and capacities could be a fruitful follow-up of the present project. Another possible move not undertaken in this paper is exploring the actual uses of the digital archives by historians of science and the ways, in which those uses are reflected in their publications.

More broadly, the study of the "past-of-science-in-the-making" could and should encompass many other dimensions: from the analysis of institutional memory policies to the affordances and constraints of the many technologies and systems that underpin the digital archive, from the particular choices made by archivists to the particular interpretations made by historians. From whichever vantage point one chooses – and certainly, there are more than those I have enumerated – the key concern remains how, through the processes of mediation and remediation, the archive is formed, and what interpretations the archive shaped in this way can lead to.

No less important, in my view, is the perspective of the archive as a field of experimentation, not only engaged with the past, but also capable of producing new understandings. This dissertation has offered some reflections in this regard. Moving forward with this “experimental” aspect of the project could involve developing indexing systems and various types of connections for different kinds of (scientific) artifacts, designing more diverse interfaces, creating ontologies capturing different interpretive approaches and developing new forms of interpretation (including machine-based ones) to understand outdated technological artifacts. And then, by experimenting with the archive and probing its capacity for generating the new, perhaps we will be able to turn the forces of “creative destruction” into a dynamic of “creative recycling”.

Bibliography

Abir-Am, Pnina G. 1982. "Essay Review: How Scientists View Their Heroes: Some Remarks on the Mechanism of Myth Construction." *Journal of the History of Biology* 15 (2): 281–315.

Abir-Am, Pnina G. 1998. *La Mise En Mémoire de La Science: Pour Une Ethnographie Historique Des Rites Commémoratifs*. Amsterdam: Éditions des Archives Contemporaines.

Abir-Am, Pnina G. 1999. "The First American and French Commemorations in Molecular Biology: From Collective Memory to Comparative History." *Osiris* 14 (1): 324–72.

Acker, Amelia. 2021. "Emulation Practices for Software Preservation in Libraries, Archives, and Museums." *Journal of the Association for Information Science and Technology* 72.

Acland, Charles R. 2012. *Swift Viewing: The Popular Life of Subliminal Influence*. Durham: Duke University Press Books.

Acland, Charles R., ed. 2007. *Residual Media*. Minneapolis: University of Minnesota Press.

Agamben, Giorgio. 2009. "What Is an Apparatus?" And Other Essays. Translated by Stefan Pedatella and David Kishik. Stanford: Stanford University Press.

Agnew, Vanessa, Jonathan Lamb, and Juliane Tomann, eds. 2019. *The Routledge Handbook of Reenactment Studies: Key Terms in the Field*. Abingdon: Routledge.

Akrich, Madeleine and Bruno Latour. 1994. "A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies." In Wiebe E. Bijker and John Law, eds. *Shaping Technology / Building Society: Studies in Sociotechnical Change*, 259–64. Cambridge, Mass.: MIT Press.

Alberti, Samuel J. M. M. 2005. "Objects and the Museum." *Isis* 96 (4): 559–71.

Alcalà, Janet Ceja, Geoffrey C. Bowker, and Susan Leigh Star. 2016. "Infrastructures for Remembering." In Geoffrey C. Bowker, Stefan Timmermans, Adele E. Clarke, and Ellen Balka, eds. *Boundary Objects and Beyond: Working with Leigh Star*, 323–38. Cambridge, Mass.: MIT Press.

Ali, Ashour, Shahrul Azman Mohd Noah, and Lailatul Qadri Zakaria. 2022. "Representation of Event-Based Ontology Models: A Comparative Study." *International Journal of Computer Science and Network Security* 22 (7): 147–56.

Allen-Robertson, James. 2018. "Critically Assessing Digital Documents: Materiality and the Interpretative Role of Software." *Information, Communication & Society* 21 (11): 1732–46.

Allen, Colin and the InPhO Group. 2013. "Cross-Cutting Categorization Schemes in the Digital Humanities." *Isis* 104 (3): 573–83.

Almeida, Mauricio Barcellos. 2013. "Revisiting Ontologies: A Necessary Clarification." *Journal of the American Society for Information Science and Technology* 64 (8): 1682–93.

Anderson, R. Joseph. 2013. "The Organization and Description of Science Archives in America." *Isis* 104 (3): 561–72.

Anderson, Steve. 2014. "Past Indiscretions: Digital Archives and Recombinant History." In Marsha Kinder and Tara McPherson, eds. *Transmedia Frictions: The Digital, the Arts, and the Humanities*, Berkley: University of California Press. 100-114.

Anheim, Étienne, and Olivier Poncet. 2004. "Fabrique des archives, fabrique de l'histoire." *Revue de Synthèse* 125 (1): 1–14.

Ankersmit, Frank R. 1988. "Historical Representation." *History and Theory* 27 (3): 205–28.

Appadurai, Arjun. 1986. "Introduction: Commodities and the Politics of Value." In Id., ed. *The Social Life of Things: Commodities in Cultural Perspective*, 3–63. Cambridge: Cambridge University Press.

Arabatzis, Theodore. 2006. *Representing Electrons: A Biographical Approach to Theoretical Entities*. Chicago: University of Chicago Press.

Arfib, Daniel. 1979. “Digital Synthesis of Complex Spectra by Means of Multiplication of Nonlinear Distorted Sine Waves.” *Journal of the Audio Engineering Society* 27 (10): 757–68.

Artières, Philippe. 2015. “L’historien face aux archives.” *Pouvoirs* 153 (2): 85–93.

Ash, Mitchell G. 1980. “Experimental Psychology in Germany before 1914: Aspects of an Academic Identity Problem.” *Psychological Research* 42 (1): 75–86.

Bachimont, Bruno. 2007. *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Paris: Hermès-Lavoisier.

Bachimont, Bruno. 2020. *Patrimoine et numérique: Technique et politique de la mémoire*. INA.

Bachimont, Bruno. 2021. “Archive et mémoire: le numérique et les mnémophores.” *Signata. Annales des sémiotiques / Annals of Semiotics* 12.

Baird, Davis. 2004. *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley: University of California Press.

Bann, Stephen. 1986. *The Clothing of Clio: A Study of the Representation of History in Nineteenth-Century Britain and France*. Cambridge: Cambridge University Press.

Bann, Stephen. 1990. *The Inventions of History: Essays on the Representation of the Past*. Manchester: Manchester University Press.

Bann, Stephen. 1994. *Under the Sign: John Bargrave as Collector, Traveler, and Witness*. Ann Arbor: The University of Michigan Press.

Barats, Christine, Valérie Schafer, and Andreas Fickers. 2020. “Fading Away... The Challenge of Sustainability in Digital Studies.” *Digital Humanities Quarterly* 14 (3).

Barritt, Marjorie Rabe. 1993. “Coming to America: Dutch *Archivistiek* and American Archival Practice.” *Archial Issues* 18 (1): 43-54.

Barthes, Roland. 1989 [1968]. “The reality effect”, In François Wahl, ed., Richard Howard, transl. *The Rustle of Language*. Berkley: University of California Press. 141-148.

Barthes, Roland. 2006 [1980]. *Camera Lucida: Reflections on Photography*. London: Vintage.

Bastian, Jeannette A. 2002. “Taking Custody, Giving Access: A Postcustodial Role for a New Century.” *Archivaria* 53: 76–93.

Bastian, Jeannette. 2016. “Moving the Margins to the Middle: Reconciling ‘the Archive’ with the Archives.” In Fiorella Foscarini, Heather MacNeil, Mak Bonnie, and Oliver Gillian, eds. *Engaging with Records and Archives: Histories and Theories*. London: Facet. 3–20.

Baumgart, Bruce G. 2019. SAILDART Prolegomenon 2018. [https://www.saildart.org/simple/book/2019.pdf]

Bearman, David A., and Richard H. Lytle. 1985. “The Power of the Principle of Provenance.” *Archivaria* 21: 14–27.

Becker, Peter, and William Clark, eds. 2001. *Little Tools of Knowledge: Historical Essays on Academic and Bureaucratic Practices*. Ann Arbor, Mich.: University of Michigan Press.

Bell, C. Gordon, and Michael M. Gold. 1972. “An Introduction to the Structure of Time-Shared Computers.” In Julius T. Tou, ed. *Advances in Information Systems Science*, Vol. 4. Boston, MA: Springer US. 161–272.

Benjamin, Walter. 1936. “L’œuvre d’art à l’époque de sa reproduction méchanisée.” In *Zeitschrift für Sozialforschung* V. 40–68.

Benschop, Ruth and Douwe Draaisma. 2000. “In Pursuit of Precision: The Calibration of Minds and Machines in Late Nineteenth-Century Psychology.” *Annals of Science* 57 (1): 1–25.

Benschop, Ruth. 1998. “What Is a Tachistoscope? Historical Explorations of an Instrument.” *Science in Context* 11 (1): 23–50.

Bergmann, Rachel. 2023. “Critical mass: collectivity and collaboration in the history of multi-agent intelligent systems.” Montreal: McGill University Press.

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. “The Semantic Web.” *Scientific American* 284 (5): 34-43.

Bert, Jean-François, and Marc Ratcliff. 2015. *Frontière d’archives: Recherches, mémoires, savoirs*. Paris: Éditions des archives contemporaines.

Bert, Jean-François. 2014. *Qu’est-ce qu’une archive de chercheur ? Qu’est-ce qu’une archive de chercheur ?* Encyclopédie numérique. Marseille: OpenEdition Press.

Bert, Jean-François. 2019. *Une histoire de la fiche érudite*. *Une histoire de la fiche érudite*. Villeurbanne: Presses de l’enssib.

Bettivia, Rhiannon S. 2016. “The Power of Imaginary Users: Designated Communities in the OAIS Reference Model.” *Proceedings of the Association for Information Science and Technology* 53 (1): 1–9.

Biagioli, Mario. 2006. “Documents of Documents: Scientists’ Names and Scientific Claims.” In Annelise Riles, ed. *Documents: Artifacts of Modern Knowledge*. Ann Arbor: University of Michigan Press. 127–57.

Biber, Douglas. 1994. “Representativeness in Corpus Design.” In Antonio Zampolli, Nicoletta Calzolari, and Martha Palmer, eds. *Current Issues in Computational Linguistics: In Honour of Don Walker*. Dordrecht: Springer. 377–407.

Biesecker, Barbara A. 2006. “Of Historicity, Rhetoric: The Archive as Scene of Invention.” *Rhetoric and Public Affairs* 9 (1): 124–31.

Blair, Ann M. 2011. *Too Much to Know: Managing Scholarly Information before the Modern Age*. New Haven: Yale University Press.

Blanchette, Jean-François. 2011. “A Material History of Bits.” *Journal of the American Society for Information Science and Technology* 62 (6): 1042–57.

Bliss, Michael. 1982: *The Discovery of Insulin*. Toronto: McClelland & Stewart.

Blom, Ina, Trond Lundemo, and Eivind Røssaak, eds. 2016. *Memory in Motion: Archives, Technology, and the Social*. Amsterdam: Amsterdam University Press.

Bloomfield, Brian P. 1987. “The Culture of Artificial Intelligence.” In Brian P. Bloomfield, ed. *The Question of Artificial Intelligence*. London: Croom Helm. 59–105.

Bloomfield, Brian P., ed. 2020. *The Question of Artificial Intelligence*. London: Routledge.

Blouin, Francis X., and William G. Rosenberg. 2011. *Processing the Past: Contesting Authority in History and the Archives*. Oxford: Oxford University Press.

Blouin, Francis Xavier, and William G. Rosenberg, eds. 2007. *Archives, Documentation, and Institutions of Social Memory*. Ann Arbor, Mich.: University of Michigan Press.

Bludau, Mark-Jan, Marian Dörk, and Frank Heidmann. 2021. “Relational Perspectives as Situated Visualizations of Art Collections.” Digital Scholarship in the Humanities 36 (Supplement 2): 17–29.

Boden, Margaret A. 1988. “Escaping from the Chinese Room.” In John Heil, ed. *Computer Models of Mind*. Cambridge: Cambridge University Press.

Boden, Margaret A. 2006. *Mind As Machine: A History of Cognitive Science*. Oxford: Oxford University Press.

Boder, D. P. 1933. “Some New Electronic Devices for the Psychological Laboratory.” *The American Journal of Psychology* 45: 145–47.

Boer, Bas de, Hedwig te Molder, and Peter-Paul Verbeek. 2021. “Understanding Science-in-the-Making by Letting Scientific Instruments Speak: From Semiotics to Postphenomenology.” *Social Studies of Science* 51 (3): 392–413.

Bogost, Ian. 2012. *Alien Phenomenology, or What It’s Like to Be a Thing*. Minneapolis: University of Minnesota Press.

Boltanski, Luc, and Arnaud Esquerre. 2014. “La «collection», une forme neuve du capitalisme la mise en valeur économique du passé et ses effets.” *Les Temps Modernes* 679 (3): 5–72.

Bolter, Jay David, and Richard Grusin. 2000. *Remediation: Understanding New Media*. Cambridge, Mass.: MIT Press.

Bondi, Marina, and Mike Scott, eds. 2010. *Keyness in Texts*. Amsterdam,: John Benjamins Publishing Company.

Bonus, Alexander. 2017. “Refashioning Rhythm: Hearing, Acting, and Reacting to Metronomic Sound in Experimental Psychology and Beyond, c.1875-1920.” In Ian D. Biddle and Kirsten Gibson, eds. *Cultural Histories of Noise, Sound and Listening in Europe, 1300-1918*. London: Routledge, Taylor & Francis. 76–106.

Boring, Edwin G. 1929. *A History of Experimental Psychology*. New York: D. Appleton-Century.

Boring, Edwin G. 1961. “The Beginning and Growth of Measurement in Psychology.” *Isis* 52 (2): 238–57.

Bosstraeten, Truus Van. 2011. “Dogs and Coca-Cola: Commemorative Practices as Part of Laboratory Culture at the Heymans Institute Ghent, 1902–1970.” *Centaurus* 53 (1): 1–30.

Boteram, Felix, and Jessica Hubrich. 2010. “Specifying Intersystem Relations: Requirements, Strategies, and Issues.” *Knowledge organization* 37 (3): 216–22.

Both, Anne. 2017. *Le sens du temps: Le quotidien d’un service d’archives départementales*. Toulouse: Anacharsis.

Boudia, Soraya, and Sébastien Soubiran. 2013. “Scientists and Their Cultural Heritage: Knowledge, Politics and Ambivalent Relationships.” *Studies in History and Philosophy of Science*, Part A 44 (4): 643–51.

Boudia, Soraya, Anne Rasmussen, Sébastien Soubiran, Jean Davallon, and François Hartog. 2010. *Patrimoine et communautés savantes*. Rennes: PU Rennes.

Bourdieu, Pierre. 1986. “L’illusion biographique.” *Actes de la Recherche en Sciences Sociales* 62 (1): 69–72.

Bowker, Geoffrey C. 2008. *Memory Practices in the Sciences*. Cambridge, Mass.: MIT Press.

Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, Mass.: MIT Press.

Brand, Stuart. 1972. “Spacewar: Fantastic Life and Symbolic Death Among the Computer Bums.” *Rolling Stone* 123: 50–58.

Breton, Ernest J. 1981. “Why Engineers Don’t Use Databases: Indexing Techniques Fail to Fit the Needs of the Profession.” *Bulletin of the American Society for Information Science* 7 (6): 20–23.

Breton, Ernest J. 1991. “Indexing for Invention.” *Journal of the American Society for Information Science* 42 (3): 173–77.

Brichford, Maynard J. 1969. *Scientific and Technological Documentation; Archival Evaluation and Processing of University Records Relating to Science and Technology*. Urbana: University of Illinois Press.

Briet, Suzanne. 2006 [1951]. *What Is Documentation?* Transl. by Ronald E. Day, Laurent Martinet, and Hermina G. B. Anghelescu. Lanham, Md.: Scarecrow Press.

Bringmann, Wolfgang G., and Gustav A. Ungerer. 1980. “The Foundation of the Institute for Experimental Psychology at Leipzig University.” *Psychological Research* 42 (1): 5–18.

Brown, Bill. 2001. “Thing Theory.” *Critical Inquiry* 28 (1): 1–22.

Brown, C. H. 1937. “The Relation of Magnitude of Galvanic Skin Responses and Resistance Levels to the Rate of Learning.” *Journal of Experimental Psychology* 20: 262–78.

Brozek, Josef, and Ludwig J. Pongratz. 1980. *Historiography of Modern Psychology*. Bern: Hogrefe & Huber.

Brügger, Niels, and Ralph Schroeder, eds. 2017. *Web as History: Using Web Archives to Understand the Past and the Present*. London: UCL Press.

Brügger, Niels. 2018. *The Archived Web: Doing History in the Digital Age*. Cambridge, Mass.: MIT Press.

Buchanan, Bruce G. 1980. “The Stanford Heuristic Programming Project: Goals and Activities.” *AI Magazine* 1 (1): 25–30.

Buchanan, Bruce G. 1983. “Introduction to the COMTEX Microfiche Edition of Memos from the Stanford University Artificial Intelligence Laboratory.” *AI Magazine* 4 (4): 37–42.

Buchwald, Jed Z. 2000. “How the Ether Spawned the Microworld.” In Lorraine Daston, ed. *Biographies of Scientific Objects*. Chicago: University of Chicago Press. 203–25.

Buckingham Shum, Simon, Enrico Motta, and John Domingue. 2000. “ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse.” *International Journal on Digital Libraries* 3 (3): 237–48.

Buckland, Michael K. 1991. “Information as Thing.” *Journal of the American Society for Information Science* 42 (5): 351–60.

Buckland, Michael K. 1997. “What Is a ‘Document’?” *Journal of the American Society for Information Science* 48 (9): 804–9.

Burton, Antoinette, ed. 2006. *Archive Stories: Facts, Fictions, And The Writing Of History*. Durham, N.C: Duke University Press.

Burton, Antoinette, ed. 2006. *Archive Stories: Facts, Fictions, And The Writing Of History*. Durham: Duke University Press.

Camras, Marvin. 1988. *Magnetic Recording Handbook*. New York: Van Nostrand Reinhold.

Canales, Jimena. 2001. “Exit the Frog, Enter the Human: Physiology and Experimental Psychology in Nineteenth-Century Astronomy.” *The British Journal for the History of Science* 34 (2): 173–97.

Canales, Jimena. 2009. *A Tenth of a Second: A History*. Chicago: University of Chicago Press.

Castelli, Lisa, Achille Felicetti, and Fabio Proietti. 2021. “Heritage Science and Cultural Heritage: Standards and Tools for Establishing Cross-Domain Data Interoperability.” *International Journal on Digital Libraries* 22 (3): 279–87.

Cattell, James McKeen, Sophie Bryant, G. F. Stout, F. Y. Edgeworth, E. P. Hughes, and C. E. Collet. 1889. “Mental Association Investigated by Experiment.” *Mind* 14 (54): 230–50.

Cattell, James McKeen. 1886a. “Psychometrische Untersuchungen, Erste Abtheilung.” *Philosophische Studien* 3: 305–35.

Cattell, James McKeen. 1886b. “The Time Taken up by Cerebral Operations. 1.” *Mind* 11 (42): 220–42.

Cattell, James McKeen. 1886c. “The Time Taken up by Cerebral Operations. 2.” *Mind* 11 (43): 377-92.

Cattell, James McKeen. 1886d. “The Time It Takes to See and Name Objects.” *Mind* 11 (41): 63–65.

Cattell, James McKeen. 1888. “The Psychological Laboratory at Leipsic.” *Mind* 13 (49): 37–51.

Caudle, Fairfid M. 1983. “The developing technology of apparatus in psychology’s early laboratories.” *Annals of the New York Academy of Sciences* 412: 19–55.

CCSDS. 2019. *Reference Model for an Open Archival Information System (OAIS). Recommended Practice*. Magenta Book. 2019.

Cerf, Vint. 1973. "PARRY Encounters the DOCTOR." Request for Comments. Internet Engineering Task Force.

Certeau, Michel de. 1992 [1975]. *The Writing of History*. Transl. by Tom Conley. New York: Columbia University Press.

Certeau, Michel de. 2011 [1980]. *The Practice of Everyday Life*. Transl. by Steven Rendall. Berkeley: University of California Press.

Chadarevian, Soraya de. 2003. "Portrait of a Discovery: Watson, Crick, and the Double Helix." *Isis* 94 (1): 90–105.

Charmasson, Thérèse. 2006. "Archives scientifiques ou archives des sciences : des sources pour l'histoire." *La revue pour l'histoire du CNRS* 14.

Chou, Siegen K. 1929. "Reaction-Keys and a New Technique for Reading-Reactions." *The American Journal of Psychology* 41 (3): 469–73.

Chun, Wendy Hui Kyong. 2008. "The Enduring Ephemeral, or the Future Is a Memory." *Critical Inquiry* 35 (1): 148–71.

Chun, Wendy Hui Kyong. 2013. *Programmed Visions: Software and Memory*. Cambridge, Mass.: MIT Press.

Clavert, Frédéric, and Muller, Caroline. 2018. *Le Goût de l'archive à l'ère numérique* [https://gout-numerique.net].

Coeckelbergh, Mark. 2020. *Using Words and Things*. London: Routledge.

Cohen, Daniel J., and Roy Rosenzweig. 2005. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press.

Cohen, Noam. 2017. *The Know-It-Alls: The Rise of Silicon Valley as a Political Powerhouse and Social Wrecking Ball*. New York: The New Press.

Colby, Kenneth Mark, Sylvia Weber, and Franklin Dennis Hilf. 1971. "Artificial Paranoia." *Artificial Intelligence* 2 (1): 1–25.

Cook, Terry. 1994. "Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era." *Archives & Manuscripts* 22 (2): 300–328.

Cook, Terry. 1997. "What Is Past Is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift." *Archivaria* 43: 17–63.

Cook, Terry. 2011. "The Archive(s) Is a Foreign Country: Historians, Archivists, and the Changing Archival Landscape." *The American Archivist* 74 (2): 600–632.

Coon, Deborah J. 1993. "Standardizing the Subject: Experimental Psychologists, Introspection, and the Quest for a Technoscientific Ideal." *Technology and Culture* 34 (4): 757–83.

Couture, Carol, and Marcel Lajeunesse. 2014. *L'archivistique à l'ère du numérique: Les éléments fondamentaux de la discipline*. Montréal: Presses de l'Université du Québec.

Crary, Jonathan. 2001. *Suspensions of Perception: Attention, Spectacle, and Modern Culture*. Cambridge, Mass.: MIT Press.

Crevier, Daniel. 1993. *AI: The Tumultuous History Of The Search For Artificial Intelligence*. New York: Basic Books, 1993.

Cunningham, Adrian. 2010. "The Postcustodial Archive." In Jennie Hill, ed. *Future of Archives and Recordkeeping: A Reader*. 173–90. London: Facet.

Dahan, Kevin. 2018. “(Re)Discovering Sounds of CCRMA – Towards Computer Music Preservation.” In *Proceedings of the 2018 International Computer Music Conference.* ICMC, Daegu, South Korea. Michigan Publishing. 134–37.

Damerow, Julia, and Dirk Wintergrün. 2019. “The Hitchhiker’s Guide to Data in the History of Science.” *Isis* 110 (3): 513–21.

Daniel, Eric D., C. Denis Mee, and Mark H. Clark. 1999. *Magnetic Recording : The First 100 Years*. New York : IEEE Press.

Danto, Arthur C. 1968. *Analytical Philosophy of History*. Cambridge: Cambridge University Press.

Danziger, Kurt. 1994. *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge: Cambridge University Press.

Daston, Lorraine, ed. 2000. *Biographies of Scientific Objects*. Chicago: University of Chicago Press.

Daston, Lorraine, ed. 2007. *Things That Talk: Object Lessons from Art and Science*. New York: Zone Books..

Daston, Lorraine, ed. 2017. *Science in the Archives: Past, Present and Future*. Chicago: University of Chicago Press.

Daston, Lorraine. 2012. “The Sciences of the Archive.” *Osiris* 27 (1): 156–87.

Davis, Natalie Zemon. 1990. *Fiction in the Archives: Pardon Tales and Their Tellers in Sixteenth-Century France*. Stanford: Stanford University Press.

Davis, Randall, Howard Shrobe, and Peter Szolovits. 1993. “What is a Knowledge Representation?” *AI Magazine* 14 (1): 17-33.

Day, Ronald E. 2008. *The Modern Invention of Information: Discourse, History, and Power*. Carbondale, Ill.: Southern Illinois University Press.

Day, Ronald E. 2014. *Indexing It All: The Subject in the Age of Documentation, Information, and Data*. Cambridge, Mass.: MIT Press.

Dearborn, Walter Fenno. 1906. *The Psychology of Reading: An Experimental Study of the Reading and Movements of the Eye*. New York: The Science press.

Debary, Octave. 2019. *De la poubelle au musée. Anthropologie des restes*. Ivry-sur-Seine: Créaphis.

Debray, Regis. 1996 [1994]. *Media Manifestos: On the Technological Transmission of Cultural Forms*. Transl. by Eric Rauth. London: Verso Books.

DeLanda, Manuel. 2016. *Assemblage Theory*. Edinburgh: Edinburgh University Press.

Deleuze, Gilles, and Claire Parnet. 1977. *Dialogues*. Paris: Flammarion.

Deloye, Juliette. 2020. [A review of] “Étienne Anheim (dir.). *Archives.* In *Annales. Histoire, sciences sociales* 74 (3-4), 2019.”

Derrida, Jacques. 1972. *La Dissémination*. Paris: Seuil.

Derrida, Jacques. 1981 [1968]. “Plato’s Pharmacy.” In Id. *Dissemination*. Ed. by Barbara Johnson. Chicago: University of Chicago Press. 61–171.

Derrida, Jacques. 1998 [1995]. *Archive Fever: A Freudian Impression*. Transl. by Eric Prenowitz. Chicago, Ill.: University of Chicago Press.

Derrida, Jacques. 1999 [1996]. “Word Processing.” *Oxford Literary Review* 21 (1): 3–17.

Derrida, Jacques. 2014 [2002]. *Trace et Archive, Image et Art*. Bry-sur-Marne: INA Éditions.

Desnoyers, Megan Floyd. 1991. “Personal Papers.” In James Gregory Bradsher, ed. *Managing Archives and Archival Institutions*. Chicago : University of Chicago Press. 78–91.

Dijck, Jose van. 2007. *Mediated Memories in the Digital Age*. Stanford: Stanford University Press.

Dionne, Claude, Silvestra Mariniello, and Walter Moser. 1996. *Recyclages : économie de l'appropriation culturelle*. Montréal : Éditions Balzac.

Dobreva, Milena, Andy O'Dwyer, and Pierluigi Feliciati, eds. 2012. *User Studies for Digital Library Development*. London: Facet.

Dodge, Raymond, and Francis Gano Benedict. 1915. *Psychological Effects of Alcohol: An Experimental Investigation of the Effects of Moderate Doses of Ethyl Alcohol on a Related Group of Neuro-Muscular Processes in Man*. Washington, D.C. : Carnegie Institution.

Doerr, Martin, and Athina Kritsotaki. 2006. "Documenting events in metadata." *The E-Volution of Information Communication Technology in Cultural Heritage*. Budapest: Archaeolingua. 56-61.

Dooley, Jackie M. 1995. "Processing and Cataloging of Archival Photograph Collections." *Visual Resources* 11 (1): 85–101.

Douglas, Mary. *How Institutions Think*. Syracuse, N.Y: Syracuse University Press, 1986.

Draaisma, Douwe. 2001 [1998]. *Metaphors of Memory: A History of Ideas about the Mind*. Transl. by Paul Vincent. Cambridge: Cambridge University Press.

Drazin, Adam. 2020. "The Object Biography." In Timothy Carroll, Antonia Walford and Shireen Walton, ed. *Lineages and Advancements in Material Culture Studies*. London: Routledge. 61-74.

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5 (1).

Drucker, Johanna. 2014. Graphesis: Visual Forms of Knowledge Production. Cambridge, Massachusetts London: Harvard University Press.

Duchein, Michel. 1983. "Theoretical Principles and Practical Problems of *Respect Des Fonds* in Archival Science." *Archivaria* 16: 64–82.

Duchein, Michel. 1992. "The History of European Archives and the Development of the Archival Profession in Europe." *The American Archivist* 55 (1): 14–25.

Duncan, Dennis. 2022. *Index, A History of the: A Bookish Adventure from Medieval Manuscripts to the Digital Age*. New York: Norton & Co.

Duranti, Luciana, and Corinne Rogers, eds. 2018. *Trusting Records and Data in the Cloud: The Creation, Management, and Preservation of Trustworthy Digital Content*. London: Facet.

Duranti, Luciana, and Patricia C. Franks, eds. 2015. *Encyclopedia of Archival Science*. Lanham: Rowman & Littlefield.

Duranti, Luciana. 1994. "The Concept of Appraisal and Archival Theory." *The American Archivist* 57 (2): 328–44.

Duranti, Luciana. 1996. "Archives as Place." *Archives & Manuscripts*, 24 (2): 242–55.

Earnest, Lester. 1980. "A Look Back at an Office of the Future." In G. P. Fick and R. H. Sprague, eds. *Decision Support Systems: Issues and Challenges*. Oxford: Pergamon Press. 119–26.

Eco, Umberto. 2009. *The Infinity of Lists: An Illustrated Essay*. Transl. by Alastair McEwen. New York: Rizzoli.

Edwards, Paul N. 1997. *The Closed World: Computers and the Politics of Discourse in Cold War America*. Cambridge, Mass.: The MIT Press.

Edwards, Paul N., Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. "Science Friction: Data, Metadata, and Collaboration:" *Social Studies of Science* 41 (5): 667–90.

Elliott, Clark A., and Pnina G. Abir-Am, eds. 2000. *Commemorative Practices in Science: Historical Perspectives on the Politics of Collective Memory*. Chicago: University of Chicago Press.

Elliott, Clark A., ed. 1983. *Understanding Progress As Process: Documentation of the History of Post-War Science and Technology in the United States*. Chicago: Society of Amer. Archivists.

Ellis, David. 1996. *Progress and Problems in Information Retrieval*. London: Library Association.

Elzway, Salem. 2023. “Armed Algorithms: Hacking the Real World in Cold War America.” *Osiris* 38: 147–64.

Erdmann, Benno, and Raymond Dodge. 1898. *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage*. Tübingen: Niemeyer.

Erll, Astrid, ed. 2012. *Mediation, Remediation, and the Dynamics of Cultural Memory*. Berlin: De Gruyter.

Ernst, Wolfgang. 1999. “Archival Action: The Archive as ROM and Its Political Instrumentalization under National Socialism.” *History of the Human Sciences* 12 (2): 13–34.

Ernst, Wolfgang. 2012 [2007]. *Digital Memory and the Archive*. Ed. by Jussi Parikka. Minneapolis: University of Minnesota Press.

Ernst, Wolfgang. 2015 [2002]. *Stirrings in the Archives: Order from Disorder*. Transl. by Adam Siegel. Lanham: Rowman & Littlefield.

Evans, Frank B., Donald F. Harrison, Edwin A. Thompson, and William L. Rofes. 1974. “A Basic Glossary for Archivists, Manuscript Curators, and Records Managers.” *The American Archivist* 37 (3): 415–33.

Evans, R. B. 2000. “Psychological instruments at the turn of the century.” *American Psychologist* 55 : 322–325.

Farge, Arlette, 1997 [1989]. *Le Goût de l’archive*. Paris: Points.

Farge, Arlette, 2013. [1989] *The Allure of the Archives*. Transl. by Thomas Scott-Railton. New Haven: Yale University Press.

Fenlon, Katrina. 2017. “Thematic Research Collections: Libraries and the Evolution of Alternative Digital Publishing in the Humanities.” *Library Trends* 65 (4): 523–39.

Fickers, Andreas, Juliane Tatarinov, and Tim van der Heijden. 2022. “Digital History and Hermeneutics – between Theory and Practice: An Introduction.” In Id., eds. *Digital History and Hermeneutics – between Theory and Practice*. De Gruyter Oldenbourg. 1–20.

Fitts, P.M. and Jones, R.E. 1947. *Analysis of factors contributing to 460 “pilot error” experiences in operating aircraft controls*. Dayton, OH: Wright-Patterson Air Force Base.

Fitts, Paul M., James R. Peterson, and Gerson Wolpe. 1963. “Cognitive Aspects of Information Processing: II. Adjustments to Stimulus Redundancy.” *Journal of Experimental Psychology* 65: 423–32.

Flanders, Julia, and Fotis Jannidis. 2015. “Knowledge Organization and Data Modeling in the Humanities.” Würzburg: Uni-Würzburg-Verlag.

Flusser, Vilém. 1999. *The Shape of Things: A Philosophy of Design*. London: Reaktion Books.

Fodor, Jerry A. 1991. “Yin and Yang in the Chinese Room.” In David Rosenthal, ed. *The Nature of Mind*. Oxford: Oxford University Press.

Folsom, Ed, 2007. “Database as Genre: The Epic Transformation of Archives.” *PMLA* 122 (5): 1571–79.

Forstner, Christian, and Mark Walker, eds. 2020. *Biographies in the History of Physics: Actors, Objects, Institutions*. Cham: Springer, 2020.

Forsythe, Diana. 2001. *Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence*. Stanford: Stanford University Press.

Foucault, Michel, 1994 [1966]. “Message Ou Bruit?” In Id. *Dits et Écrits, 1954-1988.* T. I: 1954-1969. Paris: Gallimard. 557-560.

Foucault, Michel. 1972 [1969]. *The Archaeology of Knowledge and the Discourse on Language*. Pantheon Books.

Foucault, Michel. 1980 [1969]. “What Is an Author?” In Donald F. Bouchard, ed. *Language, Counter-Memory, Practice: Selected Essays and Interviews*. Ithaca, NY: Cornell University Press. 113–39.

Foucault, Michel. 1980. “The Confession of the Flesh. Interview.” In Power/Knowledge: Selected Interviews and Other Writings, 1972-1977, 194–228. New York: Pantheon Books.

Foucault, Michel. 1994 [1966]. *The Order of Things: An Archaeology of the Human Sciences*. Reissue edition. New York: Vintage.

Franks, Patricia, ed. 2021. *The Handbook of Archival Practice*. Lanham: Rowman & Littlefield.

Fredkin, Edward. 1963. “The Time-Sharing of Computers.” *Computers and Automation* 12 (11): 12–20.

Freshwater, Helen. 2003. “The Allure of the Archive.” *Poetics Today* 24 (4): 729–58.

Friedberg, Anne. 2009. *The Virtual Window: From Alberti to Microsoft*. Cambridge, Mass.: The MIT Press.

Friedrich, Markus. 2018 [2013]. *The Birth of the Archive: A History of Knowledge*. Transl. by John Dillon. Ann Arbor: University of Michigan Press.

Frohmann, Bernd. 2004. *Deflating Information: From Science Studies to Documentation*. Toronto: University of Toronto Press.

Gabrys, Jennifer. 2011. *Digital Rubbish: A Natural History of Electronics*. Ann Arbor: University of Michigan Press.

Galloway, Alexander. 2021. *Uncomputable: Play and Politics In the Long Digital Age*. Brooklyn, NY: Verso.

Garde-Hansen, J., A. Hoskins, and A. Reading, eds. 2009. *Save As... Digital Memories*. New York: Palgrave Macmillan.

Gardner, Howard. 1985. *The Mind’s New Science: A History of the Cognitive Revolution*. New York: Basic Books.

Gartner, Richard. 2016. *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*. Cham: Springer.

Gauvin, Jean-François, Vincent Larivière, and Maxime Sainte-Marie. 2017. “Les collections muséales au service de la science. Une analyse de la Collection of Historical Scientific Instruments de l’université Harvard.” *Document numérique* 20 (2–3): 137–58.

Georges, Margot. 2016. “Institutions de recherche, chercheurs et archives des savoirs: quelle(s) définition(s) par les producteurs.” In Jean-François Bert and Marc Ratcliff, eds. *Frontières d’archives: Recherches, mémoires, savoirs*. Paris: Archives contemporaines. 29–38.

Gerber, Kent. 2023. “Digital Archives as Socially and Civically Just Public Resources.” In James O’Sullivan, ed. *The Bloomsbury Handbook to the Digital Humanities*. London: Bloomsbury Academic. 273–84.

Giaretta, David. 2011. *Advanced Digital Preservation*. Heidelberg: Springer.

Ginzburg, Carlo. 2013 [1986]. *Clues, Myths, and the Historical Method*. Transl. by John Tedeschi and Anne C. Tedeschi. Baltimore: Johns Hopkins University Press.

Gitelman, Lisa, ed. 2013. *“Raw Data” Is an Oxymoron*. Cambridge, Mass.: MIT Press.

Gitelman, Lisa. 2008. *Always Already New: Media, History, and the Data of Culture*. Cambridge, Mass.: MIT Press.

Golinski, Jan. 2008. *Making Natural Knowledge: Constructivism and the History of Science*. Chicago: University of Chicago Press.

Golub, Koraljka, and Ying-Hsang Liu, eds. 2022. *Information and Knowledge Organisation in Digital Humanities : Global Perspectives*. Taylor & Francis.

Gombrich, Ernst. H. 1994. *The Sense of Order: A Study in the Psychology of Decorative Art*. London: Phaidon Press.

Goody, Jack. 1977. *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.

Gosden, Chris, and Yvonne Marshall. 1999. “The Cultural Biography of Objects.” *World Archaeology* 31 (2): 169–78.

Got, Olivier. 2015. “«Archive(s)»: le mot.” *Sigila* 36 (2): 13–19.

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian’s Macroscope*. London: Imperial College Press.

Grange, Didier. 2016. “Quand Clio rencontre Janus. Pour une histoire de l’archivistique.” In Jean-François Bert and Marc Ratcliff, eds. *Frontière d archives. Recherches, mémoires, savoirs*. Paris: Archives contemporaines. 49–58.

Gray, Clarence Truman. 1917. *Types of Reading Ability as Exhibited through Tests and Laboratory Experiments.* Chicago: University of Chicago press.

Green, R., C. A. Bean, and Sung Hyon Myaeng. 2013. *The Semantics of Relationships: An Interdisciplinary Perspective*. New York: Springer.

Gruber, Thomas R. 1993. “A Translation Approach to Portable Ontology Specifications.” *Knowledge Acquisition* 5 (2): 199–220.

Gruber, Thomas R. 1995. “Toward Principles for the Design of Ontologies Used for Knowledge Sharing.” *International Journal of Human-Computer Studies* 43 (5–6): 907–28.

Gruman, Gerald J. 1958. “Preserving the Stuff of History.” *Science* 127 (3313): 1471–1471.

Guillory, John. 2004. “The Memo and Modernity.” *Critical Inquiry* 31 (1): 108–32.

Gundlach, Horst. 1996. “The Hipp Chronoscope as Totem Pole and the Formation of a New Tribe-Applied Psychology, Psychotechnics and Rationality.” *Teorie & Modelli* 1: 65–85.

Hackler, Ruben Marc; Kirsten, Guido. 2016. *Distant Reading, Computational Criticism, and Social Critique: an Interview with Franco Moretti*. Zürich: Foucaultblog.

Haddad, G. E. 1999. “Medicine and the Culture of Commemoration: Representing Robert Koch’s Discovery of the Tubercle Bacillus.” *Osiris* 14: 118–37.

Halbwachs, Maurice. 1980 [1950]. *The Collective Memory*. New York: Harper Colophon.

Hall, Stuart. 1999. “Un-settling ‘the Heritage’, Re-imagining the post-nationWhose Heritage?” *Third Text* 13 (49): 3–13.

Ham, F. Gerald. 1981. “Archival Strategies for the Post-Custodial Era.” *The American Archivist* 44 (3): 207–16.

Hamer, Philip M., ed. 1961. *A Guide to Archives and Manuscripts in the United States: Compiled for the National Historical Publications Commission*. New Haven: Yale University Press.

Hankins, Thomas L. 2007. “Biography and the Reward System in Science.” in Thomas Söderqvist, ed. *The History and Poetics of Scientific Biography*. Aldershot, England: Routledge. 93–104.

Harmelen, Frank van, Vladimir Lifschitz, and Bruce Porter, eds. 2008. *Handbook of Knowledge Representation*. Amsterdam: Elsevier.

Harper, Robert S. 1950. “The First Psychological Laboratory.” *Isis* 41 (2): 158–61.

Hartog, François. 2016 [2003]. *Regimes of Historicity: Presentism and Experiences of Time*. Transl. by Saskia Brown. New York: Columbia University Press.

Hartung, Michael, Toralf Kirsten, and Erhard Rahm. 2008. “Analyzing the Evolution of Life Science Ontologies and Mappings.” In Amos Bairoch, Sarah Cohen-Boulakia, and Christine Froidevaux, eds. *Data Integration in the Life Sciences*. Berlin: Springer. 11–27.

Haupt, Edward J. 1998. “Building a Cyber-Museum of Early Instruments and Procedures with the 1903 Eduard Zimmermann Catalog as an Example.” *Behavior Research Methods, Instruments, & Computers* 30 (2): 320–26.

Haupt, Edward J. 2001. "The First Memory Drum." *The American Journal of Psychology* 114 (4): 601–22.

Hawkins, Ashleigh. 2022. "Archives, Linked Data and the Digital Humanities: Increasing Access to Digitised and Born-Digital Archives via the Semantic Web." *Archival Science* 22 (3): 319–44.

Hayden, Jessica, Jane D Monson, and Emory J Trask. 2016. "Strengthening Archival Digitization Efforts with an Interdepartmental Approach: A Case Study." *Journal of Western Archives* 7 (1) 1–14.

Hayles, N. Katherine. 2002. *Writing Machines*. Cambridge, Mass.: MIT Press.

Hayles, N. Katherine. 2004. "Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis." *Poetics Today* 25 (1): 67–90.

Hayles, N. Katherine. 2005. *My Mother Was a Computer: Digital Subjects and Literary Texts*. Chicago: University of Chicago Press.

Hayles, N. Katherine. 2012. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press.

Hedstrom, Margaret, and David A. Bearman. 1993. "Reinventing Archives for Electronic Records: Alternative Service Delivery Options." In Margaret Hedstrom, ed. *Electronic Records Management Program Strategies*. Pittsburgh: Archives & Museum Informatics. 82–98.

Heesen, Anke te. 2020. "Thomas S. Kuhn, Earwitness: Interviewing and the Making of a New History of Science." *Isis* 111 (1): 86–97.

Heesen, Anke te. 2022. *Revolutionäre im Interview: Thomas Kuhn, Quantenphysik und Oral History*. Berlin: Klaus Wagenbach.

Heidegger, Martin. 1962 [1927] *Being and Time*. Transl. by J. Macquarrie, E. Robinson. Oxford: Blackwell.

Hess, Charlotte, and Elinor Ostrom, eds. 2011. *Understanding Knowledge as a Commons: From Theory to Practice*. Cambridge, Mass.: MIT Press.

Hick, W. E. 1952. "On the Rate of Gain of Information." *The Quarterly Journal of Experimental Psychology* 4 (1): 11–26.

Hill, A. R., and R. Watanabe. 1894. "'Sensorial' and 'Muscular' Reactions." *The American Journal of Psychology* 6 (2): 242–46.

Hilts, Philip J. 1982. *Scientific Temperaments: Three Lives in Contemporary Science*. New York: Simon & Shuster.

Hilts, Philip J. 1984. "Interview with John McCarthy." In Pamela Weintraub, ed. *The Omni Interview.* New York: Ticknor & Fields. 259–73.

Hine, Christine. 2006. *New Infrastructures for Knowledge Production: Understanding E-Science*. Hershey, PA: Idea Group.

Hirsch, Adolphe. 1864. "Expériences chronoscopiques sur la vitesse des différentes sensations et de la transmission nerveuse." *Bulletin de la société des sciences naturelles de Neuchâtel* 6: 100–114.

Hogan, Aidan. 2020. "The Semantic Web: Two Decades On." *Semantic Web* 11(1): 169-185.

Horner, Andrew. 2007. "Evolution in Digital Audio Technology." In Eduardo R. Miranda and John Al Biles, eds. *Evolutionary Computer Music*. London: Springer.

Hoskins, Andrew, ed. 2017. *Digital Memory Studies*. New York: Routledge.

Hoskins, Janet. 2006. "Agency, Biography and Objects." In Chris Tilley ed al., eds. *Handbook of Material Culture*. London: SAGE. 74–84.

Huey, Edmund Burke. 1908. *The Psychology and Pedagogy of Reading, with a Review of the History of Reading and Writing and of Methods, Texts, and Hygiene in Reading*. New York, MacMillan.

Huhtamo, Erkki, ed. 2011. *Media Archaeology*. Berkeley, Calif.: University of California Press.

Hui, Alexandra. 2012. *The Psychophysical Ear: Musical Experiments, Experimental Sounds, 1840-1910*. Cambridge, Mass.: MIT Press.

Hunter, Michael Cyril William. 1998. *Archives of the Scientific Revolution: The Formation and Exchange of Ideas in Seventeenth-Century Europe*. Martlesham: Boydell & Brewer.

Hunter, Nancy Chaffin, Kathleen Legg, and Beth Oehlerts. 2010. "Two Librarians, an Archivist, and 13,000 Images: Collaborating to Build a Digital Collection." *The Library Quarterly: Information, Community, Policy* 80 (1): 81–103.

Hutchby, Ian. 2001. "Technologies, Texts and Affordances." *Sociology* 35 (2): 441–56.

Ide, Nancy, and David Woolner. 2007. "Historical Ontologies." In Khurshid Ahmad, Christopher Brewster, and Mark Stevenson, eds. *Words and Intelligence II*. Dordrecht: Springer. 137–52.

Ihde, Don. 1990. *Technology and the Lifeworld: From Garden to Earth*. Bloomington: Indiana University Press.

Ihde, Don. 1991. *Instrumental Realism: The Interface between Philosophy of Science and Philosophy of Technology*. Bloomington: Indiana University Press.

Ihde, Don. 1999. *Expanding Hermeneutics: Visualism in Science*. Evanston, Ill.: Northwestern University Press.

Ilerbaig, Juan. 2016. "Organisms, Skeletons and the Archivist as Palaeontologist: Metaphors of Archival Order and Reconstruction in Context." In Fiorella Foscarini, Heather MacNeil, Mak Bonnie, and Oliver Gillian, eds. *Engaging with Records and Archives: Histories and Theories*. London: Facet. 21–40.

ISAD(G): General International Standard Archival Description. 2000. 2nd ed. ICA Standards. Ottawa: ICA.

Jacob, Christian. 2007. *Lieux de savoir.* T. 1: *Espaces et communautés*. Paris: Albin Michel.

Jansen, Ludger. 2019. "Ontologies for the Digital Humanities: Learning from the Life Sciences?" *Joint Ontology Workshops.*

Jardine, Boris. 2019. "The Museum in the Lab: Historical Practice in the Experimental Sciences at Cambridge, 1874–1936." *BJHS Themes* 4 (January): 245–71.

Jay, Martin. 2008. "Scopic Regime." In *The International Encyclopedia of Communication*. John Wiley & Sons.

Jenkinson, Sir Hilary. 1948. *The English Archivist: A New Profession*. London: H. K. Lewis.

Jensen, A. R. 1982. "Reaction Time and Psychometric g." In Hans J. Eysenck, ed. *A Model for Intelligence*. Berlin: Springer. 93–132.

Jensen, Helle Strandgaard. 2021. "Digital Archival Literacy for (All) Historians." *Media History* 27 (2): 251–65.

Johnston, John, 2010. "Technology." In W. J. T. Mitchell, Mark B. N. Hansen, eds. *Critical Terms for Media Studies*. Chicago: University of Chicago Press. 199–216.

Joy, Jody. 2009. "Reinvigorating Object Biography: Reproducing the Drama of Object Lives." *World Archaeology* 41 (4): 540–56.

Jung, Carl Gustav. 1919. *Studies in Word-Association; Experiments in the Diagnosis of Psychopathological Conditions Carried out at the Psychiatric Clinic of the University of Zurich, under the Direction of C. G. Jung.* New York: Moffat, Yard & Co.

Kahn, L. D. 1935. "An Electronic Voice-Key." *Journal of General Psychology* 12: 447–50.

Keller, Vera, Anna Marie Roos, and Elizabeth Yale, eds. 2018. *Archival Afterlives*. Leiden: Brill.

Kelty, Christopher. 2005. "Geeks, Social Imaginaries, and Recursive Publics." *Cultural Anthropology* 20 (2): 185–214.

Ketelaar, Eric. 2017. "Archival Turns and Returns." In Anne J. Gilliland, Sue McKemmish, and J. Lau, eds. *Research in the Archival Multiverse*. Clayton, Victoria: Monash University Publishing. 228–68.

Ketelaar, F. C. J. 2006. "'Everyone an Archivist.'" In N. Bütikofer, H. Hofman and S. Ross, eds. *Managing and Archiving Records in the Digital Era. Changing Professional Orientations.* Baden: hier + jetzt. 9-14.

Kirschenbaum, Matthew G. 2012. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, Mass.: MIT Press.

Kirschenbaum, Matthew G. 2016. *Track Changes: A Literary History of Word Processing*. Cambridge, Mass.: Harvard University Press.

Kirschenbaum, Matthew. 2013. "The .Txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Literary." *Digital Humanities Quarterly* 7 (1).

Kittler, Friedrich A. 1999 [1986]. *Gramophone, Film, Typewriter*. Stanford: Stanford University Press.

Knorr Cetina, Karin. 2000. "Objectual Practice." In Theodore R. Schatzki, ed. *The Practice Turn in Contemporary Theory*. London: Routledge. 175–88.

Knuth, Donald E. 1989. "The Errors of Tex." *Software: Practice and Experience* 19 (7): 607–85.

Komatsu, Lloyd. K. (1992). "Recent views of conceptual structure". *Psychological Bulletin* 112 (3): 500–526.

Kopytoff, Igor. 1986. "The Cultural Biography of Things: Commoditization as Process." In Arjun Appadurai, ed. *The Social Life of Things: Commodities in Cultural Perspective*. Cambridge: Cambridge University Press. 64–92.

Kracauer, Siegfried. 2005 [1927]. *The Mass Ornament: Weimar Essays*. Transl. by Thomas Y. Levin. Cambridge, Mass.: Harvard University Press.

Krajewski, Markus. 2011. *Paper Machines: About Cards & Catalogs, 1548-1929*. Transl. by Peter Krapp. Cambridge, Mass.: MIT Press, 2011.

Kräutli, Florian. 2016. "Visualising Cultural Data: Exploring Digital Collections Through Timeline Visualisations." Thesis, Royal College of Art.

Kuhn, Thomas Samuel. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kuhn, Thomas Samuel. 1980. "The Halt and the Blind: Philosophy and History of Science." *The British Journal for the Philosophy of Science* 31 (2): 181–92.

Kuhn, Thomas, John L. Heilbron, Paul Forman, and Lini Allen. 1967. *Sources for History of Quantum Physics: An Inventory and Report.* Philadelphia: American Philosophical Society.

LaCapra, Dominick. 1985. *History & Criticism*. Ithaca, NY: Cornell University Press.

Lamy, Jérôme, and Jean-François Bert. 2021. *Voir les savoirs - Lieux, objets et gestes de la science*. Paris: ANAMOSA.

Lancaster, F. W. 2003. *Indexing and Abstracting in Theory and Practice*. Champaign, Ill.: University of Illinois Press.

Lange, Ludwig. 1888. "Neue Experimente über den Vorgang der einfachen Reaction auf Sinneseindrücke." *Philosophische Studien* 4: 479-510.

Larsen, Knut Einar. 1994. *Architectural Preservation in Japan.* Trondheim: Tapir Publishers.

Latour, Bruno, Steve Woolgar, and Jonas Salk. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.

Latour, Bruno. 1988. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, Mass.: Harvard University Press.

Latour, Bruno. 1993. *We Have Never Been Modern*. Transl. by Catherine Porter. Cambridge, Mass.: Harvard University Press.

Latour, Bruno. 1994. "Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts." In W. E. Bijker and J. Law, eds. *Shaping Technology/Building Society*. Cambridge, Mass.: MIT Press. 225-258.

Latour, Bruno. 1995. "A door must be either open or shut: A little philosophy of techniques." In A. Feenberg & A. Hannay, eds. *Technology and the politics of knowledge*. Bloomington: Indiana University Press. 272–281.

Latour, Bruno. 2005. "From Realpolitik to Dingpolitik or How to Make Things Public." in Bruno Latour and Peter Weibel, eds. *Making Things Public: Atmospheres of Democracy*. Cambridge, Mass.: MIT Press. 14-41.

Latour, Bruno. 2007. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latour, Bruno. 2012. "Visualisation and Cognition: Drawing Things Together." *Avant: Trends in Interdisciplinary Studies* 3 (T): 207–60.

Laubichler, Manfred D., Jane Maienschein, and Jürgen Renn. 2013. "Computational Perspectives in the History of Science: To the Memory of Peter Damerow." *Isis* 104 (1): 119–30.

Laubichler, Manfred D., Jane Maienschein, and Jürgen Renn. 2019. "Computational History of Knowledge: Challenges and Opportunities." *Isis* 110 (3): 502–12.

Lavoie, Brian. 2014. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide*. DPC Technology Watch Report.

Law, John, and John Hassard. 1999. *Actor Network Theory and After*. Oxford: Wiley-Blackwell.

Law, John. 2008. "Actor Network Theory and Material Semiotics." In *The New Blackwell Companion to Social Theory.* Hoboken, NJ: John Wiley & Sons. 141–58.

Layne, Sara Shatford. 1994. "Some Issues in the Indexing of Images." *Journal of the American Society for Information Science* 45 (8): 583–88.

Le Brun, Marc. 1979. "Digital Waveshaping Synthesis." *Journal of the Audio Engineering Society* 27 (4): 250–66.

Lebo, Timothy, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. 2013. "PROV-O: The PROV Ontology." https://research.manchester.ac.uk/en/publications/prov-o-the-prov-ontology.

Lee, Christopher A. 2005. *Defining digital preservation work. A case study of the development of the reference model for an open archival information system.* Thesis. University of Michigan.

Lee, Christopher A. 2009. "Open Archival Information System (OAIS) Reference Model." In Marcia J. Bates and Mary Niles Maack, eds. *Encyclopedia of Library and Information Sciences.* CRC Press. 4020–30.

Lee, Christopher A., ed. 2011. *I, Digital: Personal Collections in the Digital Era*. Chicago, Ill.: Society of American Archivists.

Lefebvre, Muriel, and Anne-Claire Jolivet. 2020. *Des patrimoines en action: Mise en mémoire des activités scientifiques*. Toulouse: PU du Midi.

Lefebvre, Muriel, Anne-Claire Jolivet, and Sophie Dalle-Nazébi. 2016. "Les écritures ordinaires des chercheurs." In Jean-François Bert and Marc Ratcliff, eds. *Frontières d'archives: Recherches, mémoires, savoirs*. Paris: Archives contemporaines. 1–16.

Lefebvre, Muriel. 2013. "L'infra-ordinaire de la recherche. Écritures scientifiques personnelles, archives et mémoire de la recherche." *Sciences de la société* 89: 3–17.

Lepawsky, Josh. 2018. *Reassembling Rubbish: Worlding Electronic Waste*. Cambridge, Mass.: MIT Press.

Levy, David M. 1994. “Fixed or Fluid? Document Stability and New Media.” *Proceedings of the 1994 ACM European Conference on Hypermedia Technology*. New York: Association for Computing Machinery.

Levy, David M. 2011. *Scrolling Forward: Making Sense of Documents in the Digital Age*. New York: Arcade.

Levy, Steven. 2010. *Hackers: Heroes of the Computer Revolution*. Sebastopol, CA: O’Reilly.

Lewis WB. 1984. “The development of electrical counting methods in the Cavendish.” In J. Hendry, ed. *Cambridge physics in the thirties*. Bristol: Adam Hilger. 133–136.

Lindemann, Mary. 1992. “Confessions of an Archive Junkie”. In Peter Karsten and John Modell, eds. *Theory, Method and Practice in Social and Cultural History*. New York: New York University Press. 152–80.

Liu, Alan. 2018. *Friending the Past: The Sense of History in the Digital Age*. Chicago: University of Chicago Press.

Locklair, Gary. 2018. “Digital Data Demise: Obsolete Digital Data Formats.” In Mark Wolf, ed. *The Routledge Companion to Media Technology and Obsolescence*. New York: Routledge. 322–36.

Loukissas, Yanni Alexander. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. London : MIT Press.

Lourenço, Marta C., and Lydia Wilson. 2013. “Scientific Heritage: Reflections on Its Nature and New Approaches to Preservation, Study and Access.” *Studies in History and Philosophy of Science Part A* 44 (4): 744–53.

Loy, D. Gareth. 2013. “Life and Times of the Samson Box.” *Computer Music Journal* 37 (3): 26–48.

Lubar, Steven, and W. David Kingery, eds. 1995. *History From Things: Essays on Material Culture*. Washington: Smithsonian Books.

Lübbe, Hermann. 2003. *Im Zug der Zeit: Verkürzter Aufenthalt in der Gegenwart*. Berlin: Springer.

MacDonald, Isobel. 2023. “Counting When, Who and How: Visualizing the British Museum’s History of Acquisition through Collection Data, 1753–2019.” *Journal of the History of Collections* 35 (2): 305–20.

Mackenzie, Adrian. 2005. “The Performativity of Code: Software and Cultures of Circulation.” *Theory, Culture & Society* 22 (1): 71–92.

MacLeod, Roy M., and James R. Friday. 1973. “The Quest for Archives of British Men of Science.” *History of Science* 11 (1): 8–20.

Maday, Charlotte. 2016. “Contextes de recherche et conséquences sur la production documentaire des chercheurs: enjeux et resques.” In Jean-François Bert and Marc Ratcliff, eds. *Frontière d archives. Recherches, mémoires, savoirs*. Paris: Archives contemporaines. 39–46.

Malloy, Judy, ed. 2016. *Social Media Archeology and Poetics*. Cambridge, MA: MIT Press.

Manoff, Marlene. 2004. “Theories of the Archive from Across the Disciplines.” *Portal: Libraries and the Academy* 4 (1): 9–25.

Manovich, Lev, 1999. “Database as Symbolic Form.” *Convergence* 5 (2): 80–99.

Marino, Mark C. 2020. Critical Code Studies. Cambridge, Massachusetts: The MIT Press.

Markoff, John. 2006. *What the Dormouse Said: How the Sixties Counterculture Shaped the Personal Computer Industry*. New York: Penguin.

Markoff, John. 2016. Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots. New York: EccoPress.

Marvin, Carolyn. 1990. *When Old Technologies Were New: Thinking About Electric Communication in the Late Nineteenth Century*. New York: Oxford University Press.

Mauldin, Michael L. 1994. "1994-ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition." *AAAI-94 Proceedings*. 16–21.

McCarthy, John, ed. 1996. *Defending AI Research: A Collection of Essays and Reviews*. Stanford, Calif.: Center for the Study of Language and Information.

McCarthy, John, S. Boilen, Edward Fredkin, and J. C. R. Licklider. 1963. "A Time-Sharing Debugging System for a Small Computer." In E. Calvin Johnson, ed. *Proceedings of the 1963 Spring Joint Computer Conference.* Detroit, Mich. 51–57.

McCarthy, John. 1962. "Time-Sharing Computer Systems." In *Management and the Computer of the Future*. Cambridge, Mass.: MIT Press. 221–36.

McCarthy, John. 1972. "The Home Information Terminal. A 1970 View." In *Man and Computer. Proc. Int. Conf., Bordeaux 1970*. Basel: Karger. 48–57.

McCarthy, John. 1978. "History of LISP." *ACM SIGPLAN Notices* 13 (8): 217–23.

McCarthy, John. 2004. "Historical Remarks on Nonmonotonic Reasoning, Especially Circumscription." *KR'04: Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning*. Whistler, British Columbia, Canada.

McCarty, Willard. 2005. *Humanities Computing*. Basingstoke, Hampshire: Palgrave Macmillan.

McCorduck, Pamela. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. San Francisco: W.H. Freeman and company, 1978.

McGann, Jerome, 2007. "Database, Interface, and Archival Fever." *PMLA* 122 (5): 1588–92.

McGann, Jerome. 2010. "Electronic Archives and Critical Editing." *Literature Compass* 7 (2): 37–42.

McGill, Meredith L. 2007. "Remediating Whitman." *PMLA* 122 (5): 1592–96.

McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. New York: Signet.

McMeekin, Sharon M. 2011. "With a Little Help from OAIS: Starting down the Digital Curation Path." *Journal of the Society of Archivists* 32 (2): 241–53.

McPherson, M. White. 1980. The Persistence of the Apparatus of the Leipzig Laboratory in the United States: 1950–80. In W. Meischner and A. Metge, eds. Wilhelm Wundt – Progressives Erbe, Wissenschaftsentwicklung und Gegenwart. Leipzig: Karl-Marx-Universität. 164-71.

Megill, Kenneth A. 1996. *The Corporate Memory: Information Management in the Electronic Age*. London: Martindale-Hubbell.

Merkel, Julius. 1885. "Die zeitlichen Verhältnisse der Willenstätigkeit." In *Philosophische Studien* 2: 73–127.

Meroño-Peñuela, Albert, Ashkan Ashkpour, Marieke van Erp, Kees Mandemakers, Leen Breure, Andrea Scharnhorst, Stefan Schlobach, and Frank van Harmelen. 2014. "Semantic Technologies for Historical Research: A Survey." *Semantic Web* 6 (6): 539–64.

Metz, Christian. 1975. "Le signifiant imaginaire." *Communications* 23 (1): 3–55.

Meyer, David E., Allen M. Osman, David E. Irwin, and Steven Yantis. 1988. "Modern Mental Chronometry." *Biological Psychology* 26: 3–67.

Miessen, Markus, Yann Chateigné et al. 2016. *The Archive as a Productive Space of Conflict.* Berlin: Sternberg.

Milligan, Ian. 2019. *History in the Age of Abundance?: How the Web Is Transforming Historical Research*. Montreal: McGill-Queen's University Press.

Mitchell, William J. 1992. *The Reconfigured Eye: Visual Truth in the Post-Photographic Era*. Cambridge, Mass.: MIT Press.

Mooers, Calvin N. 1951. "Zatocoding Applied to Mechanical Organization of Knowledge." *American Documentation* 2 (1): 20–32.

Moretti, Franco, and Alberto Piazza. 2007. *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.

Moretti, Franco. 2000. "The Slaughterhouse of Literature." *MLQ: Modern Language Quarterly* 61 (1): 207–27.

Moretti, Franco. 2013. *Distant Reading*. London: Verso.

Morgan, Clifford T. 1963. *Human Engineering Guide to Equipment Design*. London: McGraw-Hill.

Mourat, Robin de, Ioana Ocnarescu, Anne-Lyse Renon, and Marine Royer. 2015. "Méthodologies de recherche et design : un instantané des pratiques de recherche employées au sein d'un réseau de jeunes chercheurs." *Sciences du Design* 1 (1): 68–75.

Mülberger, Annette. 2012. "Wundt Contested: The First Crisis Declaration in Psychology." *Studies in History and Philosophy of Science.* Part C: *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (2): 434–44.

Muller, S., Johan Adriaan Feith, and R. Fruin. 1968. *Manual for the Arrangement and Description of Archives*. Transl. by Arthur H. Leavitt. New York: H.W. Wilson.

Neglia, José Vicente. 2021. "'Original Artyfacts": Rethinking Object Biography in Popular Music Studies." *Volume !* 18 (2): 33–51.

Neisser, Ulric. 1967. *Cognitive Psychology*. Englewood Cliffs, N.J: Prentice Hall.

Nelson, Andrew J. 2015. *The Sound of Innovation: Stanford and the Computer Music Revolution*. MIT Press.

Nelson, Theodor H. 2009. *Geeks Bearing Gifts: How the Computer World Got This Way*. Sausalito, CA: Mindful Press.

Neville, Brian, and Johanne Villeneuve, eds. 2002. *Waste-Site Stories: The Recycling of Memory*. Albany: State University of New York Press.

Nilsson, Nils J. 2009. *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press.

Nilsson, Nils J. 2012. "John McCarthy. A Biographical Memoir." National Academy of Sciences.

Nunberg, Geoffrey. 1996. "Farewell to the Information Age." In Geoffrey Nunberg, ed. *The Future of the Book*. Berkeley: University of California Press. 103–38.

Nuttgens, Simon. 2023. "Making Psychology 'Count': On the Mathematization of Psychology." *Europe's Journal of Psychology* 19: 100–112.

O'Meara, Erin, Kate Stratton, and Society of American Archivists. 2016. *Digital Preservation Essentials*. Chicago: Society of American Archivists.

O'Neill, Brenda, and Larry Stapleton. 2022. "Digital Cultural Heritage Standards: From Silo to Semantic Web." *AI & SOCIETY* 37 (3): 891–903.

O'Shea, Geoffrey, and Theodore R. Bashore. 2012. "The Vital Role of the American Journal of Psychology in the Early and Continuing History of Mental Chronometry." *The American Journal of Psychology* 125 (4): 435–48.

Ogilvie, Brian. 2016. "Scientific Archives in the Age of Digitization." *Isis* 107 (1): 77–85.

Otlet, Paul. 1934. *Traité de documentation: le livre sur le livre, théorie et pratique*. Brussels: Mundaneum.

Otlet, Paul. 1990. "Something about Bibliography." In W. B. Rayward, ed. *International Organization and Dissemination of Knowledge: Selected Essays of Paul Otlet*. Amsterdam: Elsevier. 11–24.

Page, Ruth, and Bronwen Thomas, eds., 2011. *New Narratives: Stories and Storytelling in the Digital Age*. Lincoln: University of Nebraska Press.

Palmer, Carole L. 2004. “Thematic Research Collections.” In Susan Schreibman, Ray Siemens, and John Unsworth, eds. *A Companion to Digital Humanities*. Malden, MA: Blackwell. 348–65.

Pantalony, David. 2011. “Biography of an Artifact: The Theratron Junior and Canada’s Atomic Age.” *Scientia Canadensis* 34 (1): 51–63.

Panzarella, Giulia, Pierangelo Veltri, and Stefano Alcaro. 2023. “Using Ontologies for Life Science Text-Based Resource Organization.” *Artificial Intelligence in the Life Sciences* 3: 100059.

Parikka, Jussi. 2013. *What Is Media Archaeology?* Cambridge: John Wiley & Sons.

Peckham, Joan, and Fred Maryanski. 1988. “Semantic Data Models.” *ACM Computing Surveys* 20 (3): 153–89.

Perec, Georges. 2003. *Penser/Classer*. Paris: Seuil.

Perera, Thomas B. 1980. “Using the TRS-80 Cassette Tape Recorder as a Voice Key.” *Behavior Research Methods & Instrumentation* 12: 259–60.

Perret, Arthur. 2022. “De l’héritage épistémologique de Paul Otlet à une théorie relationnelle de l’organisation des connaissances.” These de doctorat, Bordeaux 3.

Pestre, Dominique. 1999. “Commemorative Practices at CERN: Between Physicists’ Memories and Historians’ Narratives.” *Osiris* 14: 203–16.

Peters, Isabella, and Katrin Weller. 2008. “Paradigmatic and Syntagmatic Relations in Knowledge Organization Systems.” *Information - Wissenschaft und Praxis* 59 (2): 100-7.

Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.

Pickstone, John V. 2000. *Ways of Knowing: A New History of Science, Technology and Medicine*. Manchester University Press.

Pierazzo, Elena. 2016. *Digital Scholarly Editing: Theories, Models and Methods*. London: Routledge.

Pinar Saygin, Ayse, Ilyas Cicekli, and Varol Akman. 2000. “Turing Test: 50 Years Later.” *Minds and Machines* 10 (4): 463–518.

Plato. 1998. *Phaedrus*. Transl. by James H. Nichols. Ithaca, NY: Cornell University Press.

Poncet, Olivier. 2019. “Archives et Histoire : Dépasser Les Tournants.” *Annales. Histoire, Sciences Sociales* 74 (3–4): 713–43.

Posner, Ernst. 1940. “Some Aspects of Archival Development since the French Revolution.” *The American Archivist* 3 (3): 159–72.

Posner, Michael I. 1994. *Images of Mind*. New York : Scientific American Library.

Posner, Michael I., and Ronald F. Mitchell. 1967. “Chronometric Analysis of Classification.” *Psychological Review* 74: 392–409.

Preda, Alex. 1999. “The Turn to Things: Arguments for a Sociological Theory of Things.” *The Sociological Quarterly* 40 (2): 347–66.

Preston, John, and Mark Bishop, eds. 2002. *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford: Clarendon Press.

Price, Kenneth. 2009. “Edition, Project, Database, Archive, Thematic Research Collection: What’s in a Name?” *Digital Humanities Quarterly* 3 (3).

Proctor, Robert W, and Darryl W. Schneider. 2018. “Hick’s Law for Choice Reaction Time: A Review.” *Quarterly Journal of Experimental Psychology* 71 (6): 1281–99.

Rabinbach, Anson. 1992. *The Human Motor: Energy, Fatigue, and the Origins of Modernity*. Berkeley: University of California Press.

Rajaraman, V. 2014. "JohnMcCarthy — Father of Artificial Intelligence." *Resonance* 19 (3): 198–207.

Rastier, François. 2010. "Web Semantics vs the Semantic Web?: The Problem of Keyness." *Keyness in Texts* 11: 93–110.

Rathje, William L., and Cullen Murphy. 2001. *Rubbish!: The Archaeology of Garbage*. Tucson: University of Arizona Press.

Ratto, Matt. 2011. "Critical Making: Conceptual and Material Studies in Technology and Social Life." *The Information Society* 27 (4): 252–60.

Rayward, W. Boyd. 1997. "The Origins of Information Science and the International Institute of Bibliography/International Federation for Information and Documentation (FID)." *The Journal of the Association for Information Science and Technology* 48 (4): 289-300.

Rendall, Steven, and Harold Weinrich. 2004. *Lethe: The Art and Critique of Forgetting*. Ithaca: Cornell University Press.

Renders, Hans, and Binne de Haan, eds. 2014. *Theoretical Discussions of Biography: Approaches from History, Microhistory, and Life Writing*. Leiden: Brill.

Reynolds, Lois. 2004. "Wellcome Witnesses: The Medical Research Council Applied Psychology Unit." *History of Psychology* 7: 85–93.

Rhee, Hea Lim. 2015. "Reflections on Archival User Studies." *Reference & User Services Quarterly* 54 (4): 29–42.

Rheinberger, Hans-Jörg. 1997. *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford: Stanford University Press.

Rheinberger, Hans-Jörg. 2000. "Cytoplasmic Particles. The Trajectory of a Scientific Object." In Lorraine Daston, ed. *Biographies of Scientific Objects*. Chicago: University of Chicago Press. 270–94.

Rheinberger, Hans-Jörg. 2016. "On the Possible Transformation and Vanishment of Epistemic Objects." *Teorie Vědy / Theory of Science* 38 (3): 269–78.

Richards, J. C. N., and J. P. Guilford. 1930. "A New Type of Lip Key." *Journal of Experimental Psychology* 13: 469–72.

Ricoeur, Paul. 2006 [2000]. *Memory, History, Forgetting*. Chicago: University of Chicago Press.

Rieber, R. W. 2001 [1980]. "Wundt and the Americans." In Robert W. Rieber and David K. Robinson, eds. *Wilhelm Wundt and the Making of a Scientific Psychology*. Boston: Springer. 137–51.

Riegl, Alois. 1982 [1903]. "The Modern Cult of Monuments: Its Character and Its Origin." *Oppositions* 25: 21–51.

Robertson, Craig. 2014. *Media History and the Archive*. London: Routledge.

Robertson, Craig. 2021. *The Filing Cabinet: A Vertical History of Information*. Minneapolis: University of Minnesota Press.

Robinson, David K. 2001. "Reaction-Time Experiments in Wundt's Institute and Beyond." In Robert W. Rieber and David K. Robinson, eds. *Wilhelm Wundt in History: The Making of a Scientific Psychology*. Boston: Springer. 161–204

Robinson, Edward Stevens. 1932. *Association Theory To-Day: An Essay in Systematic Psychology*. Urbana: Illinois University Press.

Romele, Alberto. 2019. *Digital Hermeneutics: Philosophical Investigations in New Media and Technologies*. New York: Routledge.

Rosenzweig, Roy. 2003. "Scarcity or Abundance? Preserving the Past in a Digital Era." *The American Historical Review* 108 (3): 735–62.

Rossiter, Margaret W. 1984. Review of *Review of Understanding Progress as Process: Documentation of the History of Post-War Science and Technology in the United States. Final Report of the Joint Committee on Archives of Science and Technology*, by Clark A. Elliott. *The American Archivist* 47 (3): 297–99.

Russell, Bertrand. 1972. *An Outline of Intellectual Rubbish: A Hilarious Catalogue of Organized and Individual Stupidity*. New York: Arno Press.

Russo Krauss, Chiara. 2017. "Back to the Origins of the Repudiation of Wundt: Oswald Külpe and Richard Avenarius." *Journal of the History of the Behavioral Sciences* 53 (1): 28–47.

Russo Krauss, Chiara. 2022. "From Introspection to Experiment: Wundt and Avenarius' Debate on the Definition of Psychology." In Davood Gozli and Jaan Valsiner, eds. *Experimental Psychology: Ambitions and Possibilities*. Cham: Springer. 7–26.

Rutherford, Ernest, C. E. Wynn-Williams, Wilfrid Bennett Lewis, and B. V. Bowden. 1933. "Analysis of α-Rays by an Annular Magnetic Field." *Proceedings of the Royal Society of London.* Series A. 139 (839): 617–37.

Samuels, Helen Willa. 1998. *Varsity Letters : Documenting Modern Colleges and Universities*. Lanham, Md.: Scarecrow Press.

Schaffer, Simon. 1988. "Astronomers Mark Time: Discipline and the Personal Equation." *Science in Context* 2 (1): 115–45.

Schellenberg, T. R. 1968. *Modern Archives, Principles and Techniques*. Chicago: University of Chicago Press.

Schmidgen, Henning, and Rand B. Evans. 2003. "The Virtual Laboratory: A New on-Line Resource for the History of Psychology." *History of Psychology* 6 (2): 208–13.

Schmidgen, Henning. 2002. "Of Frogs and Men: The Origins of Psychophysiological Time Experiments, 1850-1865." *Endeavour* 26 (4): 142–48.

Schmidgen, Henning. 2003. "Time and Noise: The Stable Surroundings of Reaction Experiments, 1860–1890." *Studies in History and Philosophy of Science.* Part C: *Studies in History and Philosophy of Biological and Biomedical Sciences* 34: 237–75.

Schmidgen, Henning. 2005. "Physics, Ballistics, and Psychology: A History of the Chronoscope in/as Context, 1845-1890." *History of Psychology* 8: 46–78.

Schmidgen, Henning. 2014. *Hirn und Zeit: Die Geschichte eines Experiments 1800-1950*. Berlin: Matthes & Seitz.

Schnapp, Jeffrey. 2018. "The Intimate Lives of Cultural Objects." In The Routledge Companion to Media Studies and Digital Humanities, 423–33. New York: Routledge.

Schulze-Kremer, Steffen, and Barry Smith. 2005. "Ontologies for the Life Sciences." In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics.* Vol. 4. Wiley.

Schuppli, Susan. 2020. *Material Witness: Media, Forensics, Evidence*. Cambridge, Mass.: MIT Press.

Schwartz, Joan M., and Terry Cook. 2002. "Archives, Records, and Power: The Making of Modern Memory." *Archival Science* 2 (1–2): 1–19.

Scott, Mike, and Christopher Tribble. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins Publishing.

Scripture, Edward W., and John Moore. 1893. "A New Reaction-key and the Time of Voluntary Movement." *Studies from the Yale Psychological Laboratory* 1: 88-90.

Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–24.

Searle, John. 2010. “Why Dualism (and Materialism) Fail to Account for Consciousness.” In Richard E. Lee, ed. *Questioning Nineteenth Century Assumptions About Knowledge*. III: *Dualism*. New York Suny Press. 5–48.

Shanks, Michael. 2011. *Experiencing the Past*. London: Routledge.

Shannon, Claude E. 1938. “A Symbolic Analysis of Relay and Switching Circuits.” *Transactions of the American Institute of Electrical Engineers* 57 (12): 713–23.

Shapin, Steven. 1984. “Pump and Circumstance: Robert Boyle’s Literary Technology.” *Social Studies of Science* 14 (4): 481–520.

Shapin, Steven. 1989.“The invisible technician.” *American Scientist*, 77 (6): 554–563.

Shapin, Steven. 1989. “Who Was Robert Hooke?” In *Robert Hooke: New Studies*, edited by Michael Hunter and Simon Schaffer, 253–85. The Boydell Press.

Shasha, Dennis, and Cathy Lazere. 1998. *Out of Their Minds: The Lives and Discoveries of 15 Great Computer Scientists*. New York: Copernicus, 1998.

Shortland, Michael, and Richard Yeo, eds. 2008. *Telling Lives in Science: Essays on Scientific Biography*. Cambridge: Cambridge University Press.

Siebold, Anna, and Matteo Valleriani. 2022. “Digital Perspectives in History.” *Histories* 2 (2): 170–77.

Sinn, Donghee, and Nicholas Soares. 2014. “Historians’ Use of Digital Archival Collections: The Web, Historical Scholarship, and Archival Research.” *Journal of the Association for Information Science and Technology* 65 (9): 1794–1809.

Smith, Barry, and Chris Welty. 2001. “Ontology: Towards a New Synthesis.” In Id., eds. *Formal Ontology in Information Systems*. New York: ACM Press.

Söderqvist, Thomas. 2007. “‘No Genre of History Fell Under More Odium than That of Biography’: The Delicate Relations between Scientific Biography and the Historiography of Science.” In Id., ed. *The History and Poetics of Scientific Biography*. Aldershot: Routledge.

Sokal, Michael M., Audrey B. Davis, and Uta C. Merzbach. 1976. “Laboratory Instruments in the History of Psychology.” *Journal of the History of the Behavioral Sciences* 12 (1): 59–64.

Soldatova, Larisa N, and Ross D King. 2006. “An Ontology of Scientific Experiments.” *Journal of the Royal Society Interface* 3 (11): 795–803.

Sowa, John F. 1983. *Conceptual Structures: Information Processing in Mind and Machine*. Reading, Mass.: Addison-Wesley.

Spieker, Sven. 2017. *The Big Archive: Art From Bureaucracy*. Cambridge, Mass.: MIT Press.

Stafford, Barbara Maria. 1996. *Artful Science: Enlightenment Entertainment and the Eclipse of Visual Education*. Cambridge, Mass.: MIT Press.

Starn, Randolph. 2002. “Authenticity and Historic Preservation: Towards an Authentic History.” *History of the Human Sciences* 15 (1): 1–16.

Steedman, Carolyn. 2002. *Dust: The Archive and Cultural History*. New Brunswick, N.J: Rutgers.

Sterling, Bruce, and Lorraine Wild. 2005. *Shaping Things*. Cambridge, Mass.: MIT Press.

Sterling, Bruce, Richard Kadrey, and Tom Jennings. 2015. *The Dead Media Notebook: 20th Anniversary Edition*. Ed. by Tom Whitwell and Music Thing Press.

Sternfeld, Joshua (2011) “Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization.” *The American Archivist*, 74: 544–575.

Stewart, Susan. 1993. *On Longing: Narratives of the Miniature, the Gigantic, the Souvenir, the Collection*. Durham: Duke University Press.

Stock, Wolfgang G. 2010. “Concepts and Semantic Relations in Information Science.” *Journal of the American Society for Information Science and Technology* 61 (10): 1951–69.

Stock, Wolfgang G. 2015. *Handbook of Information Science*. Berlin: De Gruyter Saur.

Stoler, Ann Laura. 2002. “Colonial Archives and the Arts of Governance.” *Archival Science* 2 (1): 87–109.

Stoler, Ann Laura. 2008. *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton: Princeton University Press.

Strasser, Bruno J., 2019: *Collecting Experiments: Making Big Data Biology*. Chicago: University of Chicago Press.

Strasser, Bruno. 2012. “Collecting Nature: Practices, Styles, and Narratives.” *Osiris* 27: 303–40.

Suchman, Lucy A. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press, 1987.

Sweeney, Shelley. 2008. “The Ambiguous Origins of the Archival Principle of ‘Provenance’.” *Libraries & the Cultural Record* 43 (2): 193–213.

Tagg, John. 1993. *Burden Of Representation: Essays on Photographies and Histories*. Minneapolis: University Of Minnesota Press.

Tesler, Larry. 2012. “A Personal History of Modeless Text Editing and Cut/Copy-Paste.” *Interactions* 19 (4): 70–75.

Theimer, Kate, 2012. “Archives in Context and as Context.” *Journal of Digital Humanities*, 1/2.

Thévenot, Laurent. 1984. “Rules and Implements: Investment in Forms.” *Social Science Information* 23 (1): 1–45.

Thompson, Michael. 2017. *Rubbish Theory: The Creation and Destruction of Value*. London: Pluto Press.

Thorpe, Charles, and Steven Shapin. 2000. “Who Was J. Robert Oppenheimer? Charisma and Complex Organization.” Social Studies of Science 30 (4): 545–90.

Titchener, E.B. 1905. *Experimental Psychology: A Manual of Laboratory Practice.* Vol. II. *Quantitative Experiments*. Oxford: Macmillan.

Trace, Ciaran B., and Andrew Dillon. 2012. “The Evolution of the Finding Aid in the United States: From Physical to Digital Document Genre.” *Archival Science* 12 (4): 501–19.

Turkle, Sherry, ed. 2011. *Evocative Objects: Things We Think With*. Cambridge, Mass.: MIT Press.

Turkle, Sherry. 2005. *The Second Self, Twentieth Anniversary Edition: Computers and the Human Spirit*. Cambridge, Mass.: MIT Press.

Vandendorpe, Christian, 2009: *From Papyrus to Hypertext: Toward the Universal Digital Library*. Transl. by Phyllis Aronoff and Howard Scott. Urbana: University of Illinois Press.

Vickery, B. C. 1997. “Ontologies.” *Journal of Information Science* 23 (4): 277–86.

Vieth, Ulrik. 2007. “Overview of the TEX Historic Archive”. *TUGboat* 29 (1): 73–76.

Volynskaya, Alina. 2020. “Archiving the Leftovers of Science: Metadata and Histories of Scientific Institutions.” *Digital Culture & Society* 6 (2): 133–62.

Volynskaya, Alina. 2023. “Back to Things: The Silences of the Archive and Object-Oriented Histories.” In Denise Bertschi, Julien Lafontaine Carboni, and Nitin Bathla, eds. *Unearthing Traces: Dismantling Imperialist Entanglements of Archives, Landscapes, and the Built Enviornment.* EPFL Press. 281–90.

Volynskaya, Alina. 2024. “Light Bulbs, Cricket Matches, and *Talk Softly Please*: on the Semantics of Digital Connections”. In Stefan Alschner and Katharina Günther, eds. *Sammlungsforschung im digitalen Zeitalter. Chancen, Herausforderungen und Grenzen.* Göttingen: Wallstein.

Waquet, Françoise. 2015. L'Ordre matériel du savoir. Comment les savants travaillent. XVIe-XXIe siècle. Paris: CNRS éditions.

Warnow-Blewett, Joan. 1992. "Documenting Recent Science: Progress and Needs." *Osiris* 7: 266–98.

Wells, F. L., and J. S. Rooney. 1922. "A Simple Voice Key." *Journal of Experimental Psychology* 5 (6): 419–27.

Wells, Frederic Lyman. 1908. "Normal Performance in the Tapping Test: Before and during Practice, with Special Reference to Fatigue Phenomena." *The American Journal of Psychology* 19 (4): 437–83.

Werrett, Simon. 2019. Thrifty Science: Making the Most of Materials in the History of Experiment. Chicago: University of Chicago Press.

White, Hayden. 1986. *Tropics of Discourse: Essays in Cultural Criticism*. Baltimore: Johns Hopkins University Press.

Whitelaw, Mitchell. 2015. "Generous Interfaces for Digital Cultural Collections." Digital Humanities Quarterly 9 (1).

Wilkinson, Robert T., and David Houghton. 1975. "Portable Four-Choice Reaction Time Test with Magnetic Tape Memory." *Behavior Research Methods & Instrumentation* 7 (5): 441–46.

Wise, M. Norton. 1988. "Mediating Machines." *Science in Context* 2 (1): 77–113.

Woolgar, Steve. 1985. "Why Not a Sociology of Machines? The Case of Sociology and Artificial Intelligence." *Sociology* 19 (4): 557–72.

Woolgar, Steve. 1990. "Configuring the User: The Case of Usability Trials." *The Sociological Review* 38 (1): 58–99.

Wundt, Wilhelm Max. 1874. *Grundzüge der physiologischen Psychologie*. Leipzig : Engelmann.

Yale, Elizabeth. 2015. "The History of Archives: The State of the Discipline." *Book History* 18 (1): 332–59.

Yale, Elizabeth. 2016. "The Book and the Archive in the History of Science." *Isis* 107 (1): 106–15.

Yates, Joanne. 1989. "The Emergence of the Memo as a Managerial Genre." *Management Communication Quarterly* 2 (4): 485–510.

Zaagsma, Gerben. 2023. "Digital History and the Politics of Digitization." *Digital Scholarship in the Humanities* 38 (2): 830–51.

Zattra, Laura. 2017. "Collaborating on Composition: The Role of the Musical Assistant at IRCAM, CCRMA and CSC." In Friedemann Sallis, Valentina Bertolani, Jan Burle, Laura Zattra, eds. Live Electronic Music. Composition, Performance, Study. London: Routledge.

Zimmerman, David Paul. 1991. "The Finger User Information Protocol." Internet Engineering Task Force.

Zöllner-Weber, Amélie. 2010. "Ontologies and Logic Reasoning as Tools in Humanities?" *Digital Humanities Quarterly* 3 (4).

Zundert, Joris J. van. 2016. "Screwmeneutics and Hermenumericals: The Computationality of Hermeneutics." In Susan Schreibman, Ray Siemens, and John Unsworth, eds. *A New Companion to Digital Humanities*. Hoboken, NJ: Wiley-Blackwell. 331–47.

Zundert, Joris van. 2018. "On Not Writing a Review about Mirador: Mirador, IIIF, and the Epistemological Gains of Distributed Digital Scholarly Resources." *Digital Medievalist* 11 (1).

Appendices

Table 1: Corpus of collections

Collection	Institution	Country	Institutional framework	Type of collection	If institutional: type of institution represented	Discipline represented	Genres of items exhibited	(Digital) origin
10 years on Mars	University of Michigan	USA	project	thematic		astronomy	web	native digital
Aalto Scientific Instruments Collection	Aalto University	Finland	scientific institution	scientific instruments		multiple	objects	digitized
Abdus Salam Websites Collection	The Abdus Salam Centre for Theoretical Physics	Italy	scientific institution	thematic	international research organization	multiple	web	native digital
Agricultural Systems Engineering Collection	Technische Universität München	Germany	scientific institution	thematic, technologies		life sciences	images, texts	digitized, native digital
Albert Einstein Digital Collection	Institute for Advanced Study	USA	scientific institution	personal		physics	images, texts	digitized
Alexander Graham Bell Family Papers	Library of Congress	USA	(inter)national archives	personal		engineering&technologies	texts, images	digitized
Alfred Russel Wallace Correspondence Project	Natural history museum	UK	project	personal		life sciences	texts	digitized
American Eugenics Society Records	American Philosophical Society	USA	scientific institution	institutional history	society	life sciences	texts, images	digitized
Ames Research Center Image Library	Internet Archive	USA	(inter)national archives	institutional history	research center	astronomy, engineering&technologies	images	digitized, native digital
Armour Research Foundation and IIT Research Institute	Illinois Institute of Technology	USA	scientific institution	institutional history	research institute	multiple	images	digitized, native digital
Bainbridge collection	Emilio Segré Visual Archives, American Institute of Physics	USA	scientific institution	personal		physics	images	digitized
BHL Field Notes Project	Biodiversity Heritage Library	USA	project	thematic		life sciences	texts	digitized
bitsavers.org		USA	project	thematic, technologies		computer science	software, texts, images	digitized, native digital
Caltech Digital Image Archives	California Institute of Technology	USA	scientific institution	institutional history	university	multiple	images	digitized
Canadian Think Tanks	University of Toronto	Canada	project	thematic		social sciences	web	native digital
Casebooks project	University of Cambridge	UK	project	thematic		life sciences	texts	digitized
Cavendish Laboratory collection	University of Cambridge	UK	scientific institution	institutional history	laboratory	physics	images, texts, objects	digitized
CERN Photo Archives	CERN	Switzerland	scientific institution	institutional history	international research organization	physics	images, objects	digitized
CERN video collection	CERN	Switzerland	scientific institution	institutional history	international research organization	physics	videos	digitized, native digital
Charles Babbage Center for the History of Computing	University of Minnesota	USA	history of science institute	thematic		computer science	images, texts, audio	digitized, native digital
Collection Charles Cros	Gallica BNF	France	(inter)national archives	personal		engineering&technologies	objects	digitized
Collection of scientific instruments	ETH, Zurich	Switzerland	scientific institution	scientific instruments		multiple	objects	digitized
Collection of the Society for Science and Technical Documentation	Österreichischen Mediathek	Austria	project	institutional history	society	multiple	videos	digitized
Computer Simulation and History		USA	project	thematic, technologies		computer science	software, images, texts	digitized, native digital
Coolidge, Baldwin collection	EMBL	Europe	scientific institution	personal		life sciences	images	digitized
Correspondence of Sir James Edward Smith	Linnean Society of London	UK	scientific institution	personal		life sciences	texts	digitized
Darwin's Evidence collection	American Museum of Natural History	UK, USA	project	personal, thematic		life sciences	texts	digitized
Darwin's virtual library	Biodiversity Heritage Library	USA	project	personal, thematic		life sciences	texts	digitized

Database Machine Drawings	MPIWG	Germany	scientific institution	thematic		engineering&technologies	images	digitized
Digital Collections of American Institute of Physics	Niels Bohr Library & Archives, American Institute of Physics	USA	scientific institution	institutional history, thematic	research institute	physics	texts, images, videos, audio	digitized
Digitales Archiv Mathematischer Modelle	Technische Universität Dresden	Germany	scientific institution	thematic		mathematics	objects	digitized
Discovery and Early Development of insulin	Toronto University	Canada	scientific institution, project	thematic	laboratory	life sciences	objects, texts, images	digitized
Documents Collection	Institute for Advanced Study	USA	scientific institution	institutional history	research institute	multiple	texts	digitized
Drucker archives	The Claremont Colleges	USA	scientific institution	personal		scientific management, social sciences	texts, images, videos, audio	digitized
Early Meteorology in Australia	Powerhouse museum	Australia	museum	scientific instruments		earth science	objects, images	digitized
Edward G. Mazurs Collection of Periodic Systems Images	Science History Institute	USA	history of science institute	personal, thematic		chemistry	images	digitized
Edwin A. and Marion C. Link Special Collection	Florida institute of technology	USA	scientific institution	personal		engineering&technologies	images, texts	digitized
Electropathological museum	Technisches Museum Wien	Austria	museum	institutional history	museum	life sciences	images, texts	digitized
Elliott Collection	Oxford History of Science Museum	UK	museum	personal, scientific instruments		engineering&technologies	objects	digitized
Engineering experiment station	Georgia Tech	USA	scientific institution	institutional history	laboratory	engineering&technologies	images	digitized
EPFL collection of scientific instruments	EPFL	Switzerland	scientific institution	scientific instruments		multiple	objects	digitized
Expeditions and Discoveries: Exploration and Scientific Discovery in the Modern Age	Harvard University	USA	project	thematic		multiple	images, texts	digitized
Expositions universelles	Conservatoire national des arts et métiers	France	scientific institution	thematic		scientific management, engineering&technologies	texts	digitized
Fermilab online video collection	Fermilab	USA	scientific institution	institutional history	laboratory	physics	videos	native digital
Fonds du Laboratoire Curie	Université PSL, Musée Marie Curie	France	scientific institution, museum	institutional history	laboratory	chemistry, physics	texts	digitized
Forum for Middle East Research in Anthropology	The Claremont Colleges	USA	scientific institution	thematic		social sciences	texts	digitized
Frederick Winslow Taylor Collection	Stevens Institute of Technology	USA	scientific institution	personal		scientific management	images, texts	digitized
Freeze Frame: historic polar images	Scott Polar Research Institute	Scotland	project	thematic		earth science	images	digitized
GE Research Lab Photographs	Museum of innovation and science	USA	museum	institutional history	laboratory	engineering&technologies	images, texts	digitized, native digital
Géologie collection	CNRS	France	scientific institution	thematic		earth science	images, videos	native digital
George Francis Fitzgerald Letters	Royal Dublin Society	Ireland	scientific institution	personal		physics	texts	digitized
German X-ray museum	museum-digital	Germany	museum	thematic		physics	images, texts, objects	digitized
Graz University of Technology	Technischen Universität Graz	Austria	scientific institution	institutional history	university	multiple	images, texts	digitized, native digital
Harvard Collection of Historical Scientific Instruments	Harvard University	USA	scientific institution	scientific instruments	university	multiple	objects	digitized
Hermann Joseph Muller collection	Cold Spring Harbour	USA	scientific institution, project	personal		life sciences	images, texts	digitized
Historische Instrumentensammlung des Johannes-Müller-Instituts für Physiologie	Humboldt University	Germany	scientific institution	scientific instruments, institutional history	research institute	life sciences	objects	digitized
History of FORTRAN and FORTRAN II	Computer History Museum	USA	project	thematic, technologies		computer science	software	native digital
Ingenium collection	Ingenium Museum	Canada	museum	scientific instruments, technologies		multiple	objects	digitized, native digital

Institution of Engineering and Technology (IET) Archives	London's screen archives	UK	museum	thematic		engineering&technologies	videos	digitized
Instruments for science: chemistry collection	Smithsonian Collections	USA	(inter)national archives	thematic, technologies		chemistry	texts	digitized
Instruments, Apparatus, and Other Manuals and Catalogs collectio	Cummings Center, the University of Akron	USA	history of science institute	thematic		social sciences	texts, images	digitized
IT collection	UK Web Archive	UK	project	thematic		computer science	web	native digital
it museum DataArt	Data Art	Russia	museum	thematic, technologies		computer science	objects	digitized
Jena Collection X-ray movies collection	University of Jena	Germany	project	thematic		physics, chemistry	videos	digitized
Jim Roberts Photographs	Northwestern University	USA	scientific institution	institutional history, personal	university	multiple	images	digitized
John Vincent Atanasoff papers	Iowa State University	USA	scientific institution	personal		computer science	images, texts	digitized
L'ONRSII, bureau des inventions de l'entre-deux-guerres	CNRS	France	scientific institution	institutional history	bureau des inventions	engineering&technologies	images, videos	digitized
Leiden Observatory papers	Leiden University	Netherlands	scientific institution	institutional history	observatory	astronomy	texts, images	digitized
Leo Szilard Papers	UC San Diego	USA	scientific institution	personal		physics	texts, images, videos, audio	digitized
Lick Observatory Photographical Archive (I and II)	UCSC	USA	scientific institution	institutional history	observatory	astronomy	images	digitized
Linnaean Annotated Library	Linnaean Society	UK	scientific institution	personal		life sciences	texts	digitized
Linus Pauling research notebooks	Oregon State University	USA	project	personal, thematic		life sciences, chemistry	texts, images	digitized
Louis Pasteur: travaux sur les fermentations	Institut Pasteur	France	scientific institution	thematic		life sciences	images, objects	digitized
Marconi collection	Oxford History of Science Museum	UK	museum	personal, technologies		engineering&technologies	objects	digitized
Margaret Cruikshank Papers	UCLA	USA	scientific institution	personal		social sciences	texts	digitized
Mathematics collection	University of Toronto	Canada	scientific institution	scientific instruments		mathematics	objects	digitized
Mathematics: The Winton Gallery	Science Museum Group	UK	museum	thematic		mathematics	objects	digitized
Mathématiques collection	CNRS	France	scientific institution	thematic		mathematics	images, videos	native digital
Meteorological observations	Royal society	UK	scientific institution	thematic	society	earth science	images, texts	digitized
Microscope collection	Museum Optischer Instrumente	Germany	museum	scientific instruments		multiple	objects	digitized
MIT Distinctive Collections	MIT	USA	scientific institution	institutional history	university	multiple	images, texts	digitized, native digital
Museum Collection of Physics Teaching Laboratories	Acervo Museológico dos Laboratórios de Ensino de Física	Brazil	scientific institution	scientific instruments	laboratory	physics	objects	digitized
NASA audio collection	Internet Archive	USA	(inter)national archives	thematic		astronomy	audio, video	digitized, native digital
National Museum of Science and Technology Photograph collection	National Museum of Science and Technology	Canada	museum	thematic		engineering&technologies	images	digitized
Newton's Papers	University of Cambridge	UK	scientific institution	personal		mathematics, physics	texts	digitized
Niklas Luhmann note box	Niklas Luhmann-Archiv	Germany	project	personal		social sciences	texts	digitized
NRC digital repository	National Research Council Canada	Canada	scientific institution	institutional history	research council	multiple	images	digitized, native digital
Papers of Georg and Max Bredig	Science History Institute	USA	history of science institute	personal		physics, chemistry	texts, images, objects	digitized

Photographic Collection - Miscellaneous Items	EMBL	Europe	scientific institution	institutional history	laboratory	life sciences	images	digitized
Photothèque d'Inria	National Institute for Research in Digital Science and Technology	France	scientific institution	institutional history	research institute	computer science	images, videos	native digital
Radiation safety information computational center collection	Oak Ridge National Laboratory	USA	scientific institution	thematic, technologies	research center	computer science, physics	software	native digital
Reuleaux Kinematic Mechanisms Collection	Cornell University	USA	scientific institution	scientific instruments		physics	objects	digitized
Richard A. (Dick) Rose Collection	MIT museum	USA	museum, scientific institution	scientific instruments		multiple	objects	digitized
Robert Boyle's workdiaries	Centre for Editing Lives and Letters	UK	project	personal		chemistry	texts, images	digitized
Royal Greenwich Observatory Archives	University of Cambridge	UK	scientific institution, project	institutional history, thematic	observatory	astronomy	texts	digitized
San Luis Observatory	Dudley Observatory	USA	scientific institution	institutional history	observatory	astronomy, institutional history	images, texts	digitized
Science & Mathematics collections	National Museum of American history	USA	museum	scientific instruments		multiple	objects	digitized
Science & Technology collection	National Library of Ireland	Ireland	project	thematic		multiple	web	native digital
Science Blogs Web Archive	Library of Congress	USA	project	thematic		multiple	web	native digital
Sigmund Freud Papers	Library of Congress	USA	(inter)national archives	personal		social sciences	texts, images, video	digitized
Silberrad Glassware Collection	Science Museum Group	UK	museum	thematic		chemistry	objects	digitized
Smithsonian Field Book Project	Smithsonian Collections	USA	project	thematic		multiple	texts, images	digitized
Software Documentation	CERN	Switzerland	scientific institution	thematic, technologies		computer science	software	native digital
Solar observations	e-manuscripta.ch	Switzerland	project	thematic		astronomy	texts, images	digitized
Sound & Science: Digital Histories	Humboldt University	Germany	project	thematic		multiple	audio, objects, images, texts, videos	digitized, native digital
Stanford artificial intelligence laboratory records	Stanford University	USA	scientific institution	institutional history	laboratory	computer science	texts, images, videos, audio	digitized, native digital
Thomas Edison Papers	Rutgers University	USA	project	personal		engineering&technologies	texts, images, video	digitized
Thomas Harriot online	ECHO, MPIWG	Germany	scientific institution, project	personal		astronomy	texts	digitized
TIB AV portal	Leibniz information center for science and	Germany	scientific institution	thematic		multiple	videos	digitized, native digital
Trew Collection of Letters	Friedrich-Alexander-Universität Erlangen-Nürnberg	Germany	scientific institution	personal		life sciences	texts	digitized
Turing digital archive	University of Cambridge	UK	project	personal		computer science	texts, images, videos	digitized
Universitätsarchiv	TU Berlin	Germany	scientific institution	institutional history	university	multiple	images, texts	digitized
Universitetshistorisk gjenstandsbase	University of Oslo	Norway	scientific institution	scientific instruments, institutional history	university	multiple	objects	digitized
Video Collection	American Philosophical Society	USA	scientific institution	thematic		multiple	videos	native digital
William Cooper Collection	Carnegie Mellon University	USA	scientific institution	personal		computer science, scientific management	images, texts	digitized, native digital
World's Fair Photographs Series	Henry Ford museum	USA	museum	thematic		scientific management, engineering&technologies	images	digitized
Yerkes Observatory Archives	University of Chicago	USA	scientific institution	institutional history	observatory	astronomy	images	digitized
ZCOM Zuse computer museum	museum-digital	Germany	museum	thematic, technologies		computer science	objects	digitized

Curriculum Vitae

Alina Volynskaya

E-mail: alina.volynskaya@epfl.ch
Phone: + 41 76 215 01 51
Current affiliation: Laboratory for the History of Science and Technology, Digital Humanities Institute, EPFL, Lausanne, Switzerland

Education

2020 – present	**PhD student.** Laboratory for the History of Science and Technology, Digital Humanities Institute, EPFL, Lausanne. **Thesis:** Querying the Digital Archive of Science: Distant Reading, Semantic Modelling and Representation of Knowledge. **Supervisors:** Prof. Jérôme Baudry, Dr. Jean-François Bert.
2016 –2019	**Master's degree in Digital Humanities.** University of Lausanne**.**
2011 – 2013	**Master's degree in Cultural Studies.** Department of Philosophy, Higher School of Economics, Moscow.
2007 – 2011	**Bachelor's degree in Cultural Studies**. Department of Philosophy, Higher School of Economics, Moscow.

Research experience

2020 – present	Swiss National Foundation Grant: *Querying the Digital Archive of Science: Distant Reading, Semantic Modelling and Representation of Knowledge.*
Feb – June 2019	Intern**,** *Digital Memory Project*, CERN, Geneva.
2015 – 2016	Fellowship for a short-term research stay, University of Fribourg. Project: *Modernism as a Soviet Anti-Canon: Literary Debates of the 1960-1970s*
2013 – 2014	Researcher, Moscow Institute for Social and Cultural Programs.

Teaching Experience

2019 – present	Teaching assistant, EPFL: master's course *History and the Digital.*
2013 – 2016	Lecturer, Department of Cultural Studies, Higher School of Economics, Moscow.

Publications

Volynskaya, Alina (2024) Collective Memory through Computer Memories: Retracing and Interpreting the Archive of the Stanford Artificial Intelligence Laboratory – an article to appear in the special issue of *Memory Studies Review*, ed. by Frédéric Clavert and Sarah Gensburger *(under review)*

Volynskaya, Alina (2024) Light Bulbs, Cricket Matches, and *Talk Softly Please*: on the Semantics of Digital Connections. In *Sammlungsforschung im digitalen Zeitalter. Chancen, Herausforderungen und Grenzen,* ed. By Stefan Alschner and Katharina Günther, 79-92. Göttingen: Wallstein Verlag.

Volynskaya, Alina (2023) Back to Things: the Silences of the Archive and Object-Oriented Histories. In *Unearthing Traces,* Ed. by Denise Bertschi, Julien Lafontaine Carboni, Nitin Bathla, 281-289. Lausanne: EPFL Press.

Volynskaya, Alina (2020) Archiving the Leftovers of Science: Metadata and Histories of Scientific Institutions. *Digital Culture & Society* 6(1): 133-162.

Volynskaya, Alina (2018) Modernism as a Soviet Anti-Canon: Literary Debates of the 1960-1970s. In *Logos* (6): 197-224 (*in Russian*).

Conferences

2023 Interpreting AI Memories. *Bridging AI scholarship and Memory Studies.* C2DH, University of Luxembourg.

2023 From inside the archive: modelling the history of the reaction key. *XLII Scientific Instrument Symposium.* University of Palermo.

2023 Ideal chronicler and digital leftovers: interrogating the archive of Stanford Artificial Intelligence Laboratory. *Science, Expertise and other Modes of Knowledge: Trends, Patterns, and Prospects (STS-CH 2023).* University of Basel.

2023 The slaughterhouse of science: from scientific leftovers to cultural heritage to historical data. *DARIAH Annual Event 2023, Cultural Heritage Data as Humanities Research Data?* Budapest.

2023 Digital representations: encountering the past of science through digital collections. *Is Digital Better? Collection Research in the Digital Age.* Klassik Stiftung Weimar.

2022 The slaughterhouse of science: turning scientific leftovers into historical data (poster). *Datafication in the Historical Humanities: Reconsidering Traditional Understandings of Sources and Data.* German Historical Institute, Washington DC.

2021 New order of things. Digital archives and knowledge representation. *Digital Minds - 5th Rencontres de l'EDAR. EPFL, Lausanne.*

2018 Towards an operationalization of commentary. International Conference *Translatio Formae*, University of Lausanne, The Interdisciplinary Centre for Literary Studies (CIEL). Co-authored with Michail Maiatsky.

2014 Pragmatics of Decadence: Readers' Circles and Text Reception (Story of Venus and Tannhäuser by Aubrey Beardsley). *International Conference Decadence in Europe: les Mots et les Choses*, Saint Petersburg State University.

2013 The Landscape of Ruin: Towards Aestheticization of Industrial Decay. International Conference *Between Labour and Leisure: Towards a New "Economy of Salvation"?*, Higher School of Economics, Moscow.