# Characterizing Swiss Alpine Lakes: from Wikipedia to Citizen Science

Master Thesis

**LIN Yuanhui**

Supervised by
Prof. Daniel Gatica-Perez

Social Computing Group, Idiap Research Institute
Ecole Polytechnique Fédérale de Lausanne, Switzerland
Lausanne, Switzerland
July, 2022

# Characterizing Swiss Alpine Lakes: from Wikipedia to Citizen Science

Master Thesis

**LIN Yuanhui**

## Abstract

In Switzerland, there are more than 1500 lakes located above 2000 meters of altitude. In order to understand the ecological impacts of climate change on bacteria communities in these high mountain lakes, researchers in project 2000Lakes analyze their chemistry and biology. Within the scope of this citizen science project, an educational platform has been created using the data mainly collected from Wikipedia and from the research results. By presenting Swiss alpine lakes and the 2000Lakes project in an interactive way, the goal of this platform is to raise interest and promote awareness about Swiss alpine lakes to the public, and ultimately, to conserve these ecosystems by joining forces with local citizens. Volunteers are invited to answer a survey that contains a list of questions regarding Swiss alpine lakes, whose answer can be found on the platform, and a list of usability questions, to which volunteers answer based on their own interact experience with the platform. The results of this survey are used to evaluate the effectiveness of this platform. Even though the general feedback is positive, some possible improvements are identified and hoped to be implemented in the future.

## Résumé

En Suisse, il existe plus de 1500 lacs situés à plus de 2000 mètres d'altitude. Afin de comprendre les impacts écologiques du changement climatique sur les communautés de bactéries dans ces lacs de haute montagne, les chercheurs du projet 2000Lakes analysent leur chimie et leur biologie. Dans le cadre de ce projet de science citoyenne, une plateforme éducative a été créée à partir des données principalement collectées sur Wikipédia et des résultats de la recherche. En présentant les lacs alpins suisses et le projet 2000Lakes de manière interactive, l'objectif de cette plateforme est de susciter l'intérêt du public et de le sensibiliser aux lacs alpins suisses, et finalement, de conserver ces écosystèmes en joignant les forces des citoyens locaux. Les volontaires sont invités à répondre à une enquête qui contient une liste de questions concernant les lacs alpins suisses, dont la réponse se trouve sur la plateforme, et une liste de questions d'ergonomie, auxquelles les volontaires répondent en fonction de leur propre expérience d'interaction avec la plateforme. Les résultats de cette enquête sont utilisés pour évaluer l'efficacité de cette plateforme. Même si le retour général est positif, certaines améliorations possibles sont identifiées et devraient être mises en œuvre à l'avenir.

**Keywords:** Swiss alpine lakes, climate change, microbial study, citizen science, interactive platform, web-based learning.

## Contents

## 1 INTRODUCTION

Climate change is a global issue that people are concerned about. Numerous studies have been conducted to better understand its causes and impacts, solutions are also suggested to mitigate the global warming process. From the latest assessment report released by the Intergovernmental Panel on Climate Change (IPCC) [23], for Europe specifically, current 1.1°C warmer world is already causing negative effects on natural and human systems, especially in the southern areas. For example, the structure of ecosystems in terrestrial, freshwater, and ocean have all been changed and species range shifts have happened. It also results in species extinction, sea-level rise, marine heatwaves, wildfire, and negative effects on human's well-being. The risks caused by climate change will become more severe when the global warming level gets higher, and then the adaptations taken to deal with the risks will on the contrary become less effective. Therefore, it is important that human societies do not accelerate global warming and even better, take action to ease the situation.

Lakes play an important part in this issue. They are generally considered as the sentinels of climate change because they are sensitive to and can respond rapidly to environmental forcing, especially for lakes with high altitude [20]. Not only that, as researchers summarized in [24], lakes are also integrators of the past climate change, storing information in sediments, and regulators of the future one, mainly because of their importance in global carbon cycling. To study lakes in the context of climate change, authors of [5] identified physical, chemical and biological variables of a lake that can be regarded as indicators of environmental effect. The results showed that the same variable can have different efficacy for different types of lakes in different geographical regions, and as such, different combinations of variables should be chosen depending on different research objects.

Various regional lake studies have been carried out by researchers. In [7], researchers studied the lakes in central Europe in particular, and identified many impacts of climate change on lakes such as increased water temperature and shortened ice coverage period. However, the alpine lakes in Switzerland are still poorly studied due to their sheer number and inaccessibility. The 2000Lakes project therefore aims to better understand the unexplored alpine lakes in order to monitor and assess climate change. Researchers study the chemistry and biology of these lakes and are particularly interested in analyzing the abundance of bacteria communities as a response variable, as those microorganisms are both key primary producers and organic matter degraders which drive many element cycles including those of carbon and nitrogen, and their evolution can be observed within a short time period because they undergo rapid population changes. Other parameters like water temperature, dissolved oxygen level, conductivity, pH value, etc. are also measured in order to give a thorough study on the lakes. This ongoing research aims to provide the knowledge needed to transition toward a sustainable society based on these data. Currently, over 20 lakes have been sampled out of more than 1500 Swiss alpine lakes. It is desired that general citizens get involved in this project, get research updates, and participate in the sampling process in the Swiss Alps, and hopefully, become more interested in the subject and willing to take action in daily life to contribute to environmental protection.

In the context of citizen science, a platform is hoped to be created as a complementary tool for 2000Lakes project. One of the main functions of this platform is to make users informed of this ongoing project. Users should be able to get information on what this 2000Lakes project is about, which parameters the researchers measure from the sampled lake water, the research results, and the meaning of the results on the lake ecosystem. Apart from such information, the progress of the 2000Lakes project will also be shown on this platform. It invites the public to join this scientific research by any means, and motivates the interested users to participate in the sampling campaign, which can be considered as offline crowd-sourcing activity.

However, this platform is not only about visualizing results from 2000Lakes, even though it is the primary goal. Additional data about Swiss Alpine lakes will also be valuable as they can provide a more complete image of the objects that 2000Lakes studies. By adding extra information, such as the photos people taken from the

lakes, this platform intends to present this scientific research in a more vivid way, so that the users will be more likely to get related if they have visited one of the lakes under study and grow interest to this scientific project.

It is ideal that the scientific data will not be presented in the form of dull charts which can be difficult to catch laypeople's attention and for them to comprehend. One way to make it intriguing is to make users more engaged while browsing the platform. Interactive features are therefore sought. Nevertheless, the measured scientific parameters should be well explained for non-professionals to understand.

The evaluation of this platform is also an important part of this project. Because this is a citizen science project, how well the public can get involved will be an essential criteria for assessing the platform. Having citizens use this platform and comments on it can offer valuable insights into how to improve it.

Having explained the background of 2000Lakes project and this thesis project, it then leads to the research questions this project intends to answer. The goal is to explore how scientific data should be depicted for citizen science purpose. This thesis breaks the main question into three research questions including different aspects:

- **RQ1 - Data**: Which data should be included as auxiliary information to the scientific research results?
- **RQ2 - Visualization**: How to make the platform attractive to the public?
- **RQ3 - Evaluation**: How to get valuable feedback from the users in order to improve the platform?

## 2 RELATED WORK

Literature review is mostly conducted by focusing on related projects where interactive platforms for citizen science purposes have been created. Some very famous citizen science projects such as Galaxy Zoo [21] are studied in order to learn from their success experience. The Swiss alpine lakes should be plotted on a map to show the audiences their geolocations, hence map visualization takes up an important part of this project. Special attention was paid to those projects with a map implemented on the platform.

Galaxy Zoo is a large scale citizen science project where volunteers are asked to do morphological classification on galaxy images by answering multiple-choice questions. The website makes this task easy for everyone to do by showing the graphic of each option explicitly. It invites anyone who can access to the internet to participate in this research and more than 200,000 volunteers have made their contribution. In [21], researchers examined the motivations of people taking part in this crowdsourced astronomy project. A forum survey was first set up on the Galaxy Zoo Forum and people were encouraged to answer the question "What Makes Galaxy Zoo Interesting?" freely. Then, 22 volunteers were invited to have a half-hour interview. From the responses they got, the authors classified the motivations into several categories, and found out that for this specific project at least, volunteers mostly participated because of the desire to be entertained, to learn new things, to participant in a community, and to discover something unique.

Authors of [17] also studied motivations of people joining citizen science projects. They showed that openness-to-change values, such as the desire to pursue pleasure and novelty are the main drivers for initial participation, which align with the findings in [21]. They also showed that in maintaining sustaining participation, self-transcendence values, such as the desire to protect everyone's well-being and nature, play a more important role.

These identified motivations can have prevalence in other citizen science projects. In [18] and [11] for example, researchers designed and implemented games on the platform in order to make it fun to interact with and attract users. This methodology matches with the finding that people are more likely to join citizen science projects if they can get entertained in the process.

Citizen Sort is a web-based platform for crowd-sourced labelling of various organisms. The collected labels will be used as a database for further research. In [18], the authors addressed the motivations of people participating in citizen science project through game, and analyzed how game can possibly affect the quality of collected data. In order to get the public involved and keep them stay involved in the project, the authors believed that

the platform should be as enjoyable and motivating as possible. Their final construction of the platform and evaluation of it were discussed in [19]. A series of artifacts were designed and implemented for comparison purpose, from "tool-like" with no additional motivational elements to "game-like" with entertaining storytelling. For evaluation, they invited participants to use different artifacts and give feedback. Experts, who are professional scientists in this field, enthusiasts, who have an intrinsic interest in science, and gamers, who have no interest in science but enjoy playing games, were among the participants. They found out that even though gaming elements demotivated experts and enthusiasts, gamers, who may take up a larger proportion of the public, showed more interest in this kind of implementation. However, since it will take too much effort for researchers to implement a complete game with beautiful scenes, appealing stories, and engaging tasks just for a scientific research, a good trade-off may be adopting the task gamification approach, which involves adding gaming elements to scientific tasks rather than making game the main focus. It is also worth noting that the task should be made easily understandable by simplifying the technical language.

In another project, a game called TagATune was implemented in order to gather labelled audio data from players [11]. A new mechanism was applied in this game. Instead of traditional labelling task in which each individual will be asked to choose a label for the target object, TagATune requires two players to finish the task together. The players will have access to the same or different audio clips and they are asked to describe it to one another using the tags provided. In the end, they need to determine if the clips they listen to are the same or different. By using this new mechanism, this game attracted much more participants compared to other games that collect audio metadata. TagATune is not a fancy game with lots of video game elements, but it still achieved a success of having more than 14,000 people participated within the first seven months. A good game mechanism therefore, can be essential.

When presenting 2000Lakes project on platform, it is inevitable that scientific concepts needs to be explained to laypeople so that they can better interpret the research results. The platform is therefore considered to be educational. It is ideal that a game with educational purpose can be implemented on the platform. In [12], Malone tried to answer what makes educational games fun and the findings can be valuable for this thesis project.

Malone believes there are three essential characters that can make an instructional game enjoyable, which are challenge, fantasy and curiosity [12]. In terms of challenge, a good and compelling goal should be set obvious to the player, the outcome however, should be uncertain, so that the player will not be bored from an easy straightforward game. It is also important that the players' self-esteem get boosted if they achieve the goal. In terms of fantasy, players should be able to imagine themselves with a physical object or in a social situation that is not actually presented with the help of the gaming elements. Curiosity is the motivation for people with no previous knowledge to learn new things. By combining reasonable complexity level of knowledge in the gaming environment, the goal is to deliver new knowledge to the players but at the same time not to use incomprehensible contents to frustrate them. Malone's theory was followed by a few case studies of instructive educational games, where simple games were developed to make learning more efficient and interesting. In 2021, Malone extended his work in [12], identified and recategorized more motivations for learning [13]. Interpersonal motivations were added, in which cooperation is included. It is how TagATune were made interesting. The result table of motivations can be used as a guideline when implementing an educational game.

The citizen science projects described in [10], [8] and [9] are helpful examples because those platforms were developed with map visualization.

An air quality monitoring system on a community scale was implemented in the project described in [10]. Various data including smoke images, air quality data, wind data, and crowd-sourced smell reports are presented on this website. By providing such rich information, the goal of this project is to equip citizens with the necessary scientific evidences of the air pollution situation so that they can be more confident when communicating with regulators about this issue. The main user interface of this web contains many different parts to show different kinds of data, a small map window is one of them. This map is used for wind data visualization. Bars and arrows

overlap on the map, where the height of each bar indicates the number of sensors in that area, the length of an arrow indicates the wind speed, and the direction of an arrow indicates the wind direction. The evaluation of this platform was done using Google Analytics to get statistics of the platform usage and asking users to answer a survey.

A platform called Environmental Health Channel [8] serves a similar purpose as the air quality monitoring system [10] in that various data are visualized on the website in order to empower citizens in discussing environmental health issue with the authority. Sensor data of air quality, citizen reported health symptoms and personal stories with images made up of the entire dataset. The interface of this platform mainly contains two parts with a map on the top and an interactive plot on the bottom. Clickable colored polygons representing zip code regions overlap on the map. If the user clicks on one polygon, the summarized information of corresponding zip code region will be displayed. The plot displays the statistics describing air quality or health data. User can interact by clicking on the label on the axis. A story board was also implemented on this platform so to evoke people's emotions and encourage them to take action to protect the environment.

In the project discussed in [9], a mobile application called Smell Pittsburgh was implemented with citizen science purpose. This platform is again dealing with urban air pollution issue with the goals of maintaining people's well-being and striving for long-term sustainability. The system allows citizens to report the pollution odors wherever they are, shows them the places where these odors are frequently concentrated, and visualizes odor complains from citizens in real time. Therefore, it is reasonable to develop a mobile version of platform that can be easily accessed by any portable device. On the map of this app, the odor reports are visualized using triangles. Each triangle represents one report with color indicating the smell rating. If user click on one triangle, details of the corresponding report will be shown. Circles on the map represent the air quality detected by official sensor installed in that area with color indicating the severity level of the pollution. To evaluate this app, the authors analyzed the server logs and Google Analytics events to get insights from system usage, and conducted online and paper survey to see if the platform is influential enough.

It is general knowledge that data with geolocation information will be simpler to interpret if presented on a map. In practice, as the three examples explained above, additional information is added on top of the map layer. As these projects are on a city or community scale, Google road map is used to better provide an image of where the data is associated to by showing names of the main streets and buildings.

Close attention should be paid to implement a map application. In [16], researchers evaluated the usability of different map websites including Google Maps. The guideline they proposed will be a good reference for map implementation.

Authors of [15] also compiled a list of guidelines to improve map application specifically in citizen science projects. They went through a process of implementing a prototype with map-based tasks, evaluating the platform by having citizens use the platform and give feedback, and revising the platform. The platform itself aims to support volunteers to collect and submit data about invasive species. The evaluation was done by observing user's interaction with the platform and conducting user interview in a lab environment. They collected timestamps of each user finishing each predefined task, marked down the problems each user encountered, and recorded if the user finished the task correctly. From the observation session, they noticed that web-based tasks can cause more problems than expected, including not understanding how to edit a map layer, being unable to create species location map, etc. A guideline is therefore constructed with the purpose to reduce such problems for web mapping applications. Enlightenment can be obtained from this guideline, including making the map application simple, adding fun features and help tools, and so on. A suitable evaluation method is also needed in order to better integrate user feedback.

# 3 METHODOLOGY

This section includes the methodology adopted to implement this web platform. The three aspects, namely data, visualization and evaluation will be discussed in separate sub section.

## 3.1 Data

Since this platform serves the citizen science purpose for 2000Lakes project, the data from 2000Lakes makes up an important part of the data to be used. The results from this research are stored in a sampling sheet, where the name of each sampled lake, the region of it, sampled date, altitude, measured pH value, conductivity, temperature, dissolved oxygen, 16S, 18S, g23, and other information are stored. This information offers a fresh perspective on a single lake. Using this data solely, however, can have some issues. To begin with, since 2000Lakes is an ongoing project with only 24 lakes having been sampled by now, the current data is very limited when being presented on a platform, considering there are over 1500 alpine lakes in Switzerland. It is not realistic that researchers to be able to provide complete data in a short time period as the sampling and analyzing are very time-consuming and also one of the purposes of this platform is to attract people to help with the sampling. Therefore, a database with information on more Swiss mountain lakes is needed to extend the current one. Secondly, the data from 2000Lakes are mainly about scientific analysis results of the lakes, which include many concepts like pH and 16S that may be unfamiliar to the general public with no background in chemistry or biology. Presenting only such scientific data on the platform may overwhelm the users and result in making them lose interest. Therefore, additional information that is closer to people's common sense should be added. Also, the terminologies should be well explained on the platform.

We address **RQ1** by considering various sources of data. It is desired that data for additional lakes are included, and not only that, the kinds of data that are closer to people's common sense should be included, such as length, width, and visual data.

A possible source of complementary data is the Wikipedia page of a list of mountain lakes of Switzerland. There are 217 lakes in that list, containing lakes from 15 different cantons. Name of the lake, canton of the lake, and its elevation and surface area are recorded and will be used on this platform. 168 of the lakes have a link to another Wikipedia page, which is about either the glacier or village the lake locates in, or the lake itself. From the lake's own page, some attributes listed in a table were extracted, such as maximum width, length, depth, and water volume of the lake, and were integrated into the database. We also extracted the paragraphs describing the lake itself from its own Wiki page and applied natural language processing (NLP) techniques to them. From the initial assessment, named-entity recognition (NER) and text summarization did not provide any additional information other than those indicated by attributes in the table. We, therefore, decided to use only those attributes on this platform prototype. Advanced NLP algorithms were not needed at this initial stage, however, their analysis results could be included in the future if they produce additional benefits.

Since there are only 101 lakes located above 2,000 meters in the list, all the mountain lakes are kept for visualization to give rich information. All the data were scraped using python with libraries Requests and BeautifulSoup. Out of the 217 lakes, only 4 have been sampled by 2000Lakes and therefore have research results. In the end, there are in total 237 lakes in the dataset. The geo-coordinate of each lake was added using the Google Geocoding API. Then, all of the lakes were plotted on Google Maps to make it possible to check the accuracy of the coordinates and correct the wrong ones manually. A summary of the density of the lakes in each canton is in Table 1.

In order to make the map interactive and give more visual information of the lakes, geo polygon data of each canton and of each lake were sought. Geojson files of all cantons and lakes were downloaded from Swisstopo, where all the Swiss geodata were made available to the public. The geojson file for cantons will be used as an overlay layer on the map so that users can click on one canton and get further information of the lakes inside

| Canton | Number of Lakes |
|---|---|
| Appenzell Innerrhoden | 3 |
| Bern | 22 |
| Fribourg | 2 |
| Glarus | 7 |
| Graubünden | 64 |
| Ticino | 35 |
| Jura | 1 |
| Neuchâtel | 1 |
| Nidwalden | 2 |
| Obwalden | 5 |
| Schwyz | 4 |
| St. Gallen | 8 |
| Uri | 9 |
| Valais | 69 |
| Vaud | 7 |

Table 1. Number of Recorded Lakes in Each Canton

that canton. Since the file for lakes contains not only mountain lakes but also lakes located in the city, which are not the targets of this project, it was filtered by the list of lakes in the final dataset. 163 out of 237 lakes have such geo polygon information.

Images of each lake are also considered as desired visual elements. We downloaded all the images from each lake's individual Wikimedia page whenever it exists and obtained over four thousands images in total. However, not all the lakes have their own Wikimedia page and therefore some of them have no corresponding image information. In addition to this, some downloaded images dated a long time ago and do not present the current appearance of the lake (there are some black-and-white pictures). Therefore, other sources of images were needed. We compared the photos from AllTrails, a platform records hiking routes in many countries, and Google Maps. The photos from these two platforms are all uploaded by users. Because a larger population uses Google Maps, the photos from there are more complete, meaning more lakes can be found with photos. We therefore choose to use photos from Google Maps. In practice, photos are retrieved using Google Maps Places API, which allows non-commercial usage with attribution properly displayed.

To sum up, multi-sourced data are used on this platform for citizen science purpose. The core data is from 2000Lakes, but got extended using the data mostly from Wikipedia, Swisstopo, and Google API. For each lake, there are not only technical measurements like pH value and conductivity value but also measurements that are more approachable like max width and length. Visual elements of the lake will also be displayed on the platform. However, not all lakes have every attribute and this final dataset does not include all the alpine lakes in Switzerland. A complete lake dataset is desired for future work.

## 3.2 Website Visualization

To address **RQ2**, attractive and reasonable color composition need to be applied on the website. Effort should be paid to make the layout clear and interactive functions should be implemented to engage the users.

*3.2.1 Aesthetic Design.* The aesthetic design of this website is based entirely on the design of 2000Lakes logo as shown in Figure 1. The font family is chosen to be Montserrat, and the four colors in the color palette are applied

in different elements throughout the website. Aside from that, the red color with hex code #DA291C is also used as it is the color of the Swiss flag in the logo. The goal is to achieve a coherent visual tone for this platform and 2000Lakes project. There are several reasons for doing so. The key reason is that this logo will be used on the platform to indicate its close relationship with 2000Lakes project explicitly. By showing the same logo and applying the same style, it reinforces people's impression of this research project. Additionally, it is preferable to reuse the colors from the logo and avoid using redundant colors which may make the platform less visually pleasant. Last but not least, the color green and blue can be easily associated with the natural environment and fit the context well.

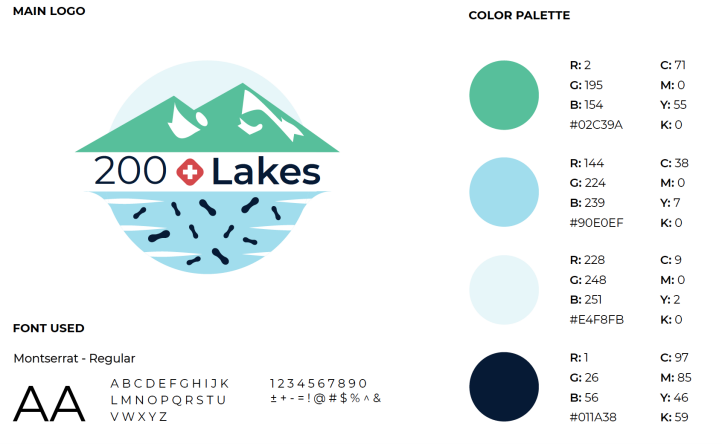



Fig. 1. Design of 2000Lakes logo.

*3.2.2 Function Design.* There are certain desired functions on this platform. To represent the current status of the 2000Lakes project, a dynamic progress bar should be displayed on the platform. By showing users the current progress of 2000Lakes, it conveys the message that this research project is still at its early stage and users' participation in any activities initiated by 2000Lakes will actually advance the research progress further. Ideally, it encourages people with interest to take action to join the research.

On the platform, the analysis results of each lake will be shown, including 16S, 18S, water temperature, dissolved oxygen, conductivity, and pH value. A section of explanation on these measured scientific parameters is required, as the target audiences may have limited knowledge of these concepts. Each parameter should be described in detail, including what it is, why it is important in the lake ecosystem, and how human activities can possibly alter it. It's critical that these parameters are not explained in academic terms.

Aside from the parameter itself, the actual measured value of each parameter should be explained as well because a pure number carries no meaning for laypeople. The exact number will not be shown to the audience, instead, the measured value will be divided into several ranges and shown on a scale. For dissolved oxygen, conductivity, and pH value, there are references on which range of value is optimal, fairly good, or terrible for the lake ecosystem [1–3], therefore, the scale of these parameters will be divided based on corresponding criteria. The scales should also be made interactive. When users click on one range, the meaning of that range will show up. By engaging users more, this activity hopes to make the knowledge more memorable. The temperature scale will be divided evenly from 0°C to 20°C, as there are no criteria on which range of temperature is good or not. The same reason applies to 16S and 18S, which represent the abundance of bacteria and archaea, and the abundance of

eukaryotic DNA respectively. However, since the value of these two parameters varies a lot from zero to billions, it is not possible to visualize the scale linearly and divide it evenly, these two parameters will be displayed in order of magnitude.

To show a lake's location, it should be plotted on a map, which, in this project, is designed with different layers. The basic one is implemented using cantons' geojson file only, colored polygons are shown with no other information. The color of each canton indicates the density of recorded lakes within that canton. The darker the color, the higher the density. Users can also choose to overlap the polygons on top of Google maps. In this case, they get to know the surrounding environment of the lake, like the mountain, the glacier, the main road, or the city. It will be helpful to locate the lake if users have existing knowledge of that area.

The map can be zoomed in or out by clicking on one canton, and the lakes inside the canton will only appear after the map is zoomed in to the canton level. This hierarchical visualization invites users to browse and interact with the map. Different visualizations will be available for lakes. Users can choose to visualize lakes based on the coordinates, the elevation, or the surface area depending on their own interests. For example, it is easier to find fun fact like the largest mountain lake in a certain canton if the user chooses to visualize lakes by surface area.

It is designed such that when a mouse hovers over a canton or a lake on the map, the basic information of the target will be displayed, eliminating the need for the user to click each time. But if the user wants further information about the lake, he or she has to click on the lake on the map. Additional data will then be displayed on the platform in another section, to which the user will be automatically directed. There should be an entry that directs the users to the lake's Wikipedia page in case they grow interest in that particular lake and would like to learn more. For lakes with no Wikipedia, the users will be directed to an empty page where they will be invited to create a Wikipedia page for that lake based on the information provided on the platform. It is hoped that by enlisting the help of citizens, the Wikipedia data on Swiss mountain lakes would be completed.

Another feature should be implemented on this platform in order to attract non-professionals is a game. Inspired by [12], where many educational games that differ from fancy video games were presented, a game similar to 10-question game will be developed. The traditional 10-question game requires two players. One player comes up with a target in mind, the other has to ask at most 10 yes-or-no questions regarding the target, and guess what the target is based on the answers given. On this platform, however, a chatbot will be built and play the game with the user. The chatbot randomly draws a lake from the dataset, instead of letting the user type the questions, it generates five multiple-choice questions, lets the user select one choice, and answers if the choice is correct or not. The first question is always about the canton the lake locates in and the rest four questions will be selected randomly from the pre-defined question pool, where the questions are all about the lake's attributes, such as max depth, measured conductivity, and so on (see Appendix B). Since the purpose of this game is to make the users learn from the platform and revise what they have learnt through browsing, the correct answer will be displayed after each question if they get it wrong. It is essential to show the player that he or she has won the game at the end of each round, so to boost the self-esteem of the player.

*3.2.3 Implementation.* Python scripts were written to preprocess the collected data and construct them in a way that is more suitable for use on the website. HTML, CSS, and JavaScript were used to implement the website. The whole implementation process also benefits from several external packages and APIs. Bootstrap was used to refine the layout, d3 library was used to visualize data and implement various interactive elements, canvas-confetti was used to support the effect after the player has won the game, and Google Maps API was used to support map functions and display photos of the lakes.

The final website prototype is presented in this section. The user interface of the first section is shown in Figure 2, it is the first look this platform appears to users once it is loaded. The bar on top is the dynamic progress bar showing the status of 2000Lakes. When this page is first loaded, the green bar will extend from a length of zero to current length, and a note stating the number of lakes sampled versus the number of lakes

recorded will be displayed. Below is a slider where the scientific parameters are explained. The five pages include microbial abundance (16S and 18S), temperature, dissolved oxygen, conductivity, and pH value. The explanation on microbial abundance was provided by one of the researchers working on 2000Lakes project. For water temperature, dissolved oxygen, conductivity, and pH value, the explanations were drafted based on [4], [2], [1], and [3] respectively, and the correctness was verified by the researcher. Users can click on the radio button below to turn the page. There will be an information icon on some pages to hint at how users can interact with the platform.

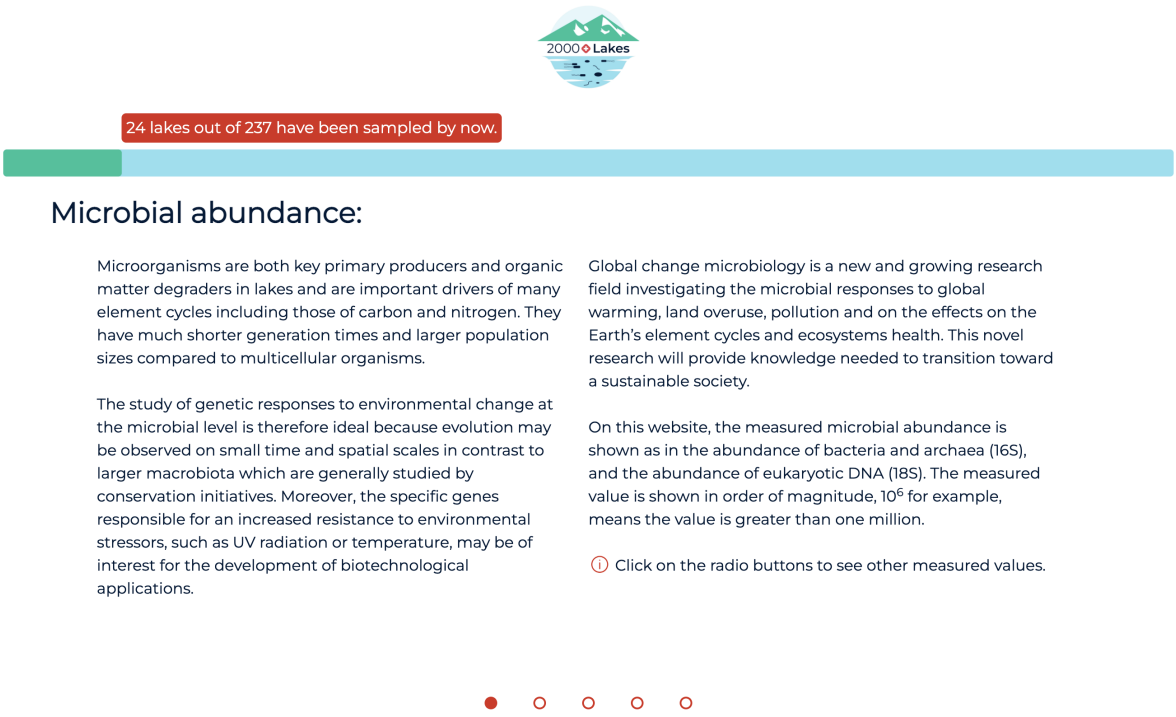



Fig. 2. First section of the website: parameter explanation.

For some of the measured parameters, namely, dissolved oxygen, conductivity, and pH value, there are criteria on how to set the range of values based on the lake ecosystem's response to the measured values [1–3]. Therefore, for the page describing the corresponding parameter, there is an explanation on how the values are divided into ranges. Interactive scales are implemented in order to make users more involved in the learning process. As shown in Figure 3, the scale on the left is the original one before the user's interaction and the one on the right is the scale after the interaction. When the user first arrives at the page, the scale is shown as gray and black, with no descriptions. But if he or she clicks on one range, the color of that range will change, and accordingly, a short text describing the meaning of that range will appear. There are different colors applied on different ranges which match people's general perception of the color scheme in that red means dangerous, values falling into this range cause the death of the organisms inside the lake, yellow means warning, values in this range cause pressure for lake organisms to live, light green means good, and dark green means ideal, the lake ecosystem will be under a good condition when the value is inside these ranges. It is hoped that the scales can help laypeople

better memorize the parameters in a visual way, so that even if they do not read the explanation carefully, they can still grasp the meaning of the measured value. The exact scales will be shown in the lake description section as well (Figure 6), using the same element is hoped to boost users' impression.

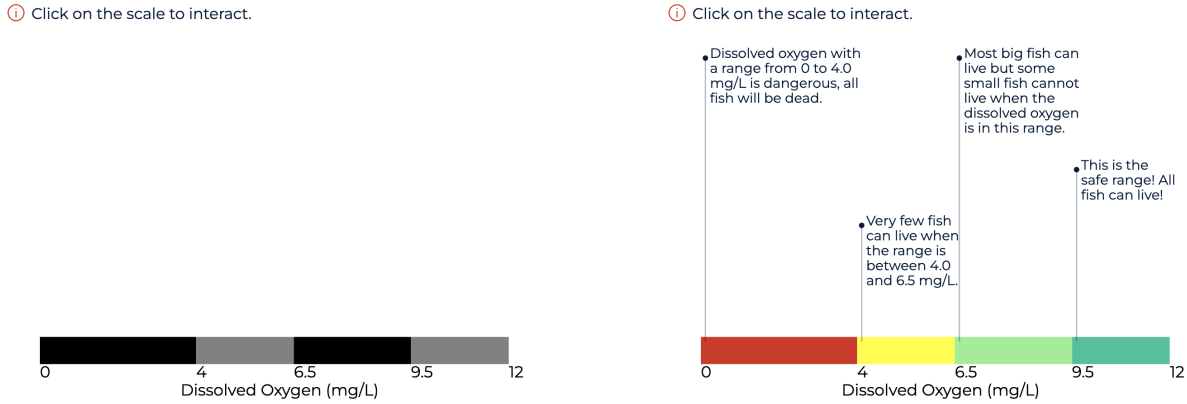



Fig. 3. Interactive scale. On the left is the default appearance of the scale. After interaction, the scale shows as on the right.

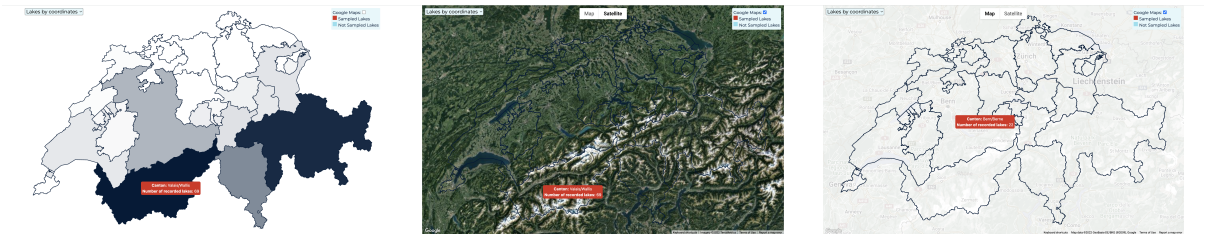


Fig. 4. Second section of the website: different map layers.

Figure 4 shows the map visualizations on this platform. There are different layers of the map serving different purposes. The default one is the left-most one showing only the colored polygons of all Swiss cantons. The darker the color, the more lakes there are. With this visualization, users get to see roughly the lake distribution in Switzerland, free from distractions. If the user checks the Google Maps option on the top-right part of the page, the map as shown in the middle of Figure 4 will appear. The default Google Maps is set to be constructed by satellite tiles because it is more suitable for mountain lakes, which are located in the natural environment. Mountains and glaciers can be easily located using satellite map, it therefore helps users perceive the lakes' elevation. Some of the Google Maps' inherent functions are preserved so that users can change the Google Maps to the road map, where the main city and main road will be shown (Figure 4, right). As citizens are usually more familiar with these information, this visualization can give a better idea of where the lake is providing their previous knowledge of the canton.

For each type of the map, when the mouse hovers over a canton, the name of the canton and its lake density will be shown in a red tooltip, moving with the mouse. At the same time, the cursor will change to a zoom-in icon to let the user know it is possible to click and zoom in to that canton. When the user chooses to zoom in to

one canton, the map will adjust the center and zoom level based on the selected canton so to display only that canton on the map.

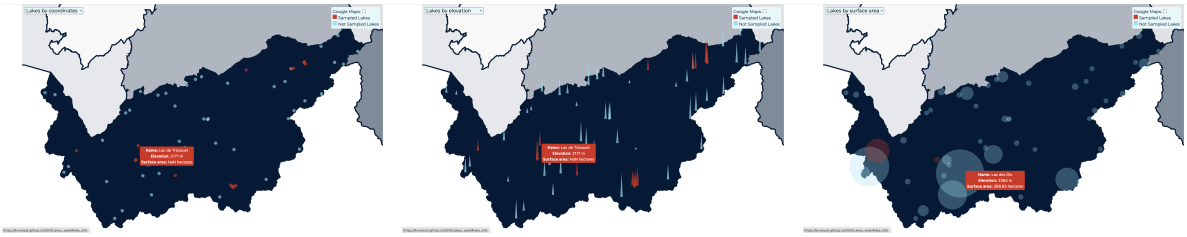


Fig. 5. Second section of the website: different lake visualizations.

Figure 5 shows different lake visualizations on map after zoomed in. Users can choose to visualize the lakes by coordinates, by elevation, or by surface area from the drop-down box on the top left part of the page, which corresponds to the left, middle and right image in Figure 5 respectively. Since not all lakes have surface area information, when they are visualized by surface area, there will be some missing points. When the mouse hovers over a lake, basic information including its name, its elevation, and its surface area will be shown in the red tooltip. The cursor will change to a zoom-out icon to prompt the user to click to zoom out. As indicated in the top-right legend, lakes displayed with red color are the ones that have been sampled and blue are the ones to be sampled.

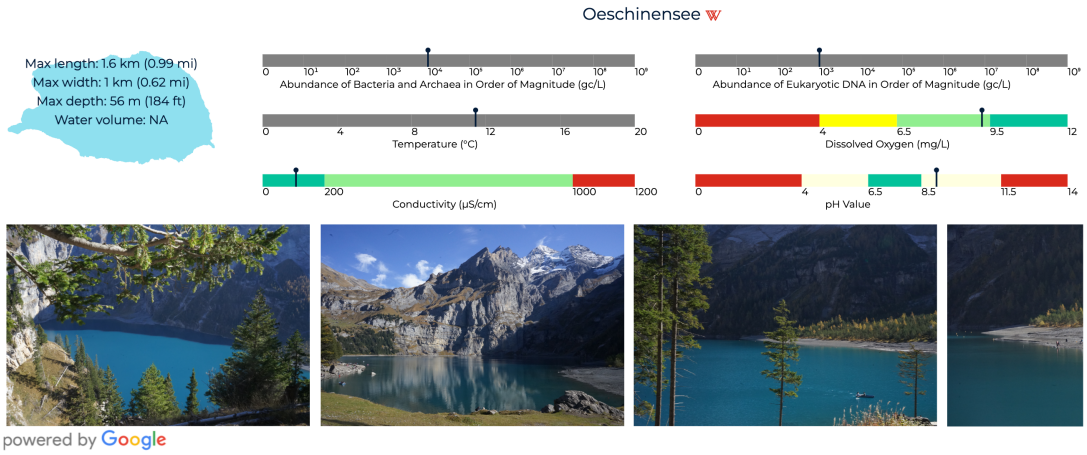



Fig. 6. Third section of the website: detailed lake information.

When the user clicks on one lake on the map, the website will redirect the user to another section where the detailed information about the lake will be shown as in Figure 6. The name of the lake is displayed in the middle with a red Wikipedia icon next to it. Users can click on the icon, if the Wiki page for this lake exists, that page will be opened in another tab in the browser, and if the page does not exist, a page written "Hey you find a lake without Wiki page! Why not create one for it!" will be shown to the user.

The lake geojson data and the data scraped from the lake's own Wiki entry are shown on top left. By showing the shape of the lake, the max depth, max width, max length, and water volume of the lake, this part intends to give the user an overview of the target. The data from 2000Lakes are visualized with the six scales. When a sampled lake is selected, the pointers will move alone the scales. If an unsampled lake is selected, however, all pointers will move back to the initial position. As explained in the first section, 16S and 18S will be shown in order of magnitude. Since there are no criteria on how to set the range for 16S, 18S, and temperature, those scales are all in gray. The rest three scales are exactly the same as in the first section.

Last but not least, a small photo gallery of the selected lake will be shown on the bottom. Photos are from Google Maps API. A maximum of 10 pictures will be shown because of Google's restriction. This part is to give a more realistic view of the lakes. And it is also the intention of this part to attract citizens with beautiful lake scenes, so that they may be more willing to join the sampling campaign in order to enjoy the beautiful nature and more likely to take other actions to safeguard the environment in order to preserve the current situation.

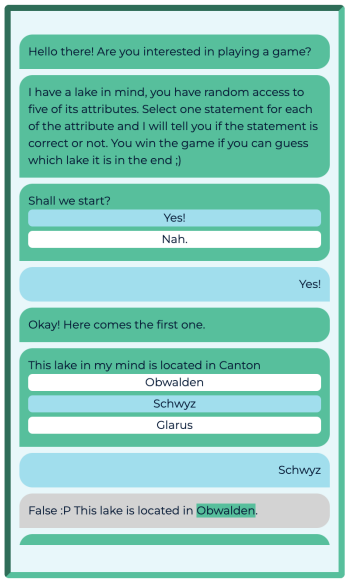



Fig. 7. Fourth section of the website: game implementation.

The implemented game is shown in Figure 7. It is designed in a way like a bot is texting with the user in a live chat box to make the experience more immersive. When the user scrolls down to this game section for the first time, the messages will be popping up into the chat box, starting with a greeting, followed by an introduction to this game, and then the user gets to choose if he or she wants to play this game or not. If the user selects "Nah.", the bot will end the conversation and ask the user to reload the page if he or she decides to play in the end. If the user chooses to play, the bot will start the game by giving out questions.

There are five questions to be guessed by the user. The first one is always about the lake's location, and the rest four are selected randomly from the question pool so that the user will not be bored by the same questions. Missing data for some lakes results in some questions bearing no useful information. In such cases, the located canton of the target lake can be very essential for players to guess the right answer and therefore this question is always asked.

Once the user makes a guess, the bot will answer immediately if the guess is right or wrong and reply accordingly. For a wrong guess, the reply bubble will be set to gray color instead of green, so that the user can distinguish the results more easily. Unlike the 10-question game, which is for amusement only, this game is designed to be informative. Therefore, even if the user makes a wrong guess, the right answer will always be revealed. In the reply message, the right answer will be highlighted using green color so that the user can easily locate that attribute of the lake. And when the user has to make a final guess of the target lake, it is easier to go back to previous dialogue and collect information from there.

In order to make the dialogue resemble people's daily communication and boost the enjoyment of interacting with this chatbot, the messages sent by the bot are all written in a friendly and encouraging way. Several predefined responses are written for replying right and wrong guess respectively, and the bot will randomly pick one to send the message. The responses for right guess include "Yes, you are correct about this!", "Congret, it's a right statement!", "Brilliant! You are right!", "Ohh you get it right!", "Yes! You are close to the final answer!", and "Correct! Keep it up!". And the responses for wrong guess include "No, you miss the right answer :(", "It's not right...", "No, it is not the case.", "Sadly it is a wrong statement.", and "False :P". The bot is humanized by using relaxing tone and avoiding fixed reply.

If the user guesses the right lake the chatbot selected, a confetti effect will be triggered as shown in Figure 8, and the bot will reply "Bravo! Are you a mind reader?!". It is hoped to make the player feel satisfied and pleased through this positive feedback so to encourage the player to continue learning about Swiss mountain lakes on this platform and through this game. If the player guesses a wrong answer, the response message will appear as gray color and the correct lake will be highlighted using green color. So that the player can match the given attributes to the target lake easily.

After one round of game is finished, the bot will continue to ask if the player wants to play again, and the cycle keeps going on.


Fig. 8. Fourth section of the website: confetti effect triggered.

### 3.3 Platform Evaluation

Current website is a working prototype which will eventually be refined. Feedback from the actual users is sought for improvement.

To answer **RQ3**, the platform is going to be evaluated from different perspectives. As it is an educational platform, where the users should be able to learn general knowledge of Swiss alpine lakes and 2000Lakes, which involves some scientific concepts. It is of interest that how well citizens with no previous knowledge are able to grasp those information. In addition, the website will be evaluated in terms of user interface and user experience design, the goal is to make the platform easy and pleasing to use. Finally, in order to make it a successful citizen science project, how influential this website is should also be evaluated.

An online survey, which is consisted of different parts, is constructed for the evaluation purpose. Participants will first read a short introduction to this project and sign a consent form in order to continue the survey. They are informed that they can stop the survey whenever they want so that participants will only finish the survey because of their own will. After this, participants will be asked to provide some general information about themselves such as how concerned they are to the environment, how often they go hiking in Switzerland, and how much they know about the Swiss alpine lakes. Having these information before participants actually interact with the platform is essential as they can be used to help decide the influence and usability of the platform. For example, if a person with a low degree of environmental concern becomes willing to take actions to preserve the environment or join 2000Lakes project after visiting the website, that is a positive feedback. And if a person with not much existing knowledge of Swiss alpine lakes can still find it easy to interact with and find information from the map, that is also a positive feedback.

The general information part is followed by a task section. In this section, ten questions are asked, whose answer can be found on the website. The questions include:

- **Q1**: How many lakes are recorded in total?
- **Q2**: For the conductivity parameter, is the higher the better?
- **Q3**: Which range of pH value is the most suitable for all kinds of fish?
- **Q4**: Which Canton has the most lakes?
- **Q5**: Lac de Salanfe in Canton Valais is in a very bad condition for the organisms inside in terms of dissolved oxygen. Is this statement true?
- **Q6**: The largest mountain lake in Canton Graubünden is ...
- **Q7**: The highest mountain lake in Canton Vaud is ...
- **Q8**: The water volume for Limmernsee in Canton Glarus is ...
- **Q9**: Guggisee in Canton Valais has a lower water temperature than Oeschinensee in Canton Bern. Is this statement true?
- **Q10**: How much does the game makes you want to learn more about mountain lakes in Switzerland?

They are designed such that participants have to go through each element on the website in order to find all the answers. It makes sure the participants actually interact with the website and get to know all the basic features so that their response on the usability evaluation part will be more reliable. Based on the correctness of these questions, it is possible to assess how good the platform is in conveying knowledge. Also, the survey will record the time spent on each question, therefore, it is possible to filter out any response that is filled in randomly.

For each question in the task section, there will also be a brief instruction on how the user can find the answer on the website. Take a multiple-choice question as an example, Q3 asks the most suitable pH value for all kinds of fish, and the instruction is written "Read the fifth slide of the measured parameters.". So, what the participant needs to do is to click on the fifth radio button as shown in Figure 2, read the passage about pH value, and click on the ranges of the scale to find the right answer. The instruction is not very detailed because the participants are also expected to explore the website by themselves.

Most questions are multiple-choice questions and yes-or-no questions, where a statement is written and the participant needs to find related information to see if the statement is right or not. To evaluate the game however, the participant is asked to play the game for as least one round, and scale how much this game makes him or her wants to learn more about Swiss alpine lakes (Q10).

After the participant finish all the tasks, some questions about the platform's aesthetics design and the whole interaction experience are asked.

To evaluate visual aesthetics, participants are asked to which degree they agree that everything goes together on this website, the layout is pleasantly varied, the color composition is attractive, and the layout appears professionally designed. Those evaluation items are in the shortened version of Visual Aesthetics of Websites inventory (VisAWI) and were proved to be reliable and able to provide a close representation of the full-length version [14].

To evaluate users' subjective impression on their experience with this website, an optimized version of user experience questionnaire (UEQ) is adopted, where the 26 items are reduced to 8 [22]. Participants are asked to scale the platform in terms of supportive versus obstructive, easy versus complicated, efficient versus inefficient, clear versus confusing, exciting versus boring, interesting versus not interesting, inventive versus conventional, and innovative versus usual.

In the end, we are interested in how influential this platform is. We access this based on the outcomes and possible actions participants may take after visiting this website. They are asked to scale to which degree they agree that they have learnt something new from the website and would like to learn more on this topic. Participants will also be asked if they become more aware of how people can affect the lakes and more concerned about the need for environmental protection. And if they would like to participate the project by helping sample lake water in real life, by creating Wikipedia page for lakes, and share the website to others.

To have a primary result for analysing, we carried out a non-remunerated survey using LimeSurvey and hoped to reach around 20 responses. This online survey was approved by Idiap's DREC Committee. Participants are mostly students in Swiss university. They received the survey link and finished the survey using their own laptop. Therefore, the testing environment varies from case to case. However, this should not be a problem, as the authors of [6] indicated that different combinations of device and software have only reasonable effect on the accuracy and precision of web platform's display and response timing. The survey was conducted in an anonymous way and only the timestamped answers will be recorded. The full survey can be found in Appendix C.

## 4 RESULTS & DISCUSSION

The majority of the survey's recipients are university students in Switzerland, who were invited to fill in the survey voluntarily. In the end, there are 26 persons started the survey but not all of them completed the whole survey. The evaluation of the platform is based on the 19 responses that are finished.

This whole survey is expected to be finished in 20 minutes including all the tasks and evaluation questions for someone who has never used the platform before. In practice, the time spent on the survey varies for different participants, this could be due to different devices and the internet connection situations, as the experiment is not conducted in a constant lab environment. The average time they spent on this survey is 17 min 23 sec, the median time is 14 min 56 sec. All the response time seem to be reasonable, therefore no response is filtered out.

Among the 19 participants, 10 are women and 9 are men. 13 of them are in the age group of 18-24 and 6 of them are in the age group of 25-34. For most of the participants, they declared that they care about the environment, with 3 of them declared to be very concerned about this issue. Therefore, this platform may be of interest of these people and they may be prone to take actions to join this citizen science project afterwards. Also, most of them go hiking in Switzerland but less than 6 times between April and September, and they do not have much knowledge about Swiss alpine lakes. That means they do not have enough previous knowledge in order to answer

| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Time | 54.05 | 58.29 | 36.92 | 32.26 | 100.51 | 61.67 | 40.94 | 51.23 | 123.42 | 64.37 |

Table 2. Median time spent on each question. Recorded in second.

all the questions about Swiss alpine lakes correctly, the correctness of their answer will be mostly based on the information collected on this platform.

### 4.1 Task Results

There are 10 tasks for participants to finish, they need to follow the brief instruction and find the related information to answer corresponding question. With the exception of the final task, where the participants are asked to rate how much the implemented game makes them want to learn more about Swiss alpine lakes, each of the remaining 9 questions has a correct answer that participants are suppose to choose. The accuracy of the response is generally very high. 15 out of 19 participants answered the questions with at most 1 mistake, and that include participants with only very little knowledge on this topic. In order to access how easy the platform is to use, we examine the time spent to answer each question. The median time spent on each question can be find in Table 2. Q10 requires user to play the game for at least one round, where the player have to interact with the chatbot, the consumed time is therefore the longer the better, meaning they indeed play the game.

Q9 is the most time-consuming question. The question is "Guggisee in Canton Valais has a lower water temperature than Oeschinensee in Canton Bern. Is this statement true?". In order to answer this questionthe participant needs to find Guggisee in Canton Valais, click on it and get detailed information on the measured temperature, and then find Oeschinensee in Canton Bern, get its temperature and finally compare the two values. This can be time-consuming because there are many lakes in Canton Valais, if the participant has no idea where Guggisee locates, he or she has to check each red points one by one in order to find the right one. The second time-consuming question is Q5, where participants are asked to answer if the condition of Lac de Salanfe in Canton Valais is bad or not in terms of dissolved oxygen. Participants need to find Lac de Salanfe, click on it and check its dissolved oxygen value. And if they are uncertain about the value's meaning, they also need to go back to the first section where this parameter is explained and compare with the values there. Again, this requires looking for one specific lake in Canton Valais, where many lakes have been sampled, so it takes more time to find the target lake. The time spent on the other questions are reasonable, participants are able to answer within one minute. We therefore assume that it is mainly the searching part, which is designed for this survey specifically, causes the time loss.

It is also worth mentioning that Q2 and Q3 are very similar questions. Q2 asks if increased conductivity is better and Q3 asks the most suitable pH value for all kinds of fish. They both instruct the users to go to the first section, read the explanation on corresponding parameter and engage with the scale. After interacting with the first conductivity scale, users become more familiar with that feature of the platform and spend less time on Q3.

The accuracy of each question is shown in Table 3. All of the participants answer Q3 and Q8 correctly. From Q2 to Q3, the spent time becomes less and the accuracy becomes higher, probably because users have adapted to this platform. Q8 asks the water volume for Limmernsee in Canton Glarus. Since there are not many recorded lakes in Canton Glarus, it is easier to find the target lake and read the water volume information from the detailed lake page.

Q9 is the question that gets the most wrong answers and at the same time it is the most time-consuming one. The question asks participants to compare two lakes. One possibility of this low accuracy is that participants get impatient finding the target lakes and decide to answer randomly.

| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 |
|---|---|---|---|---|---|---|---|---|---|
| Accuracy | 84.21% | 94.74% | 100% | 94.74% | 89.47% | 78.95% | 94.74% | 100% | 68.42% |

Table 3. Accuracy of each question.

Another question that gets relatively low accuracy is the first question, where it asks the number of recorded lakes in this platform. The participants are asked to load and stay in the first section of the page and the recorded number can be found above the progress bar. The correct answer is 237, but some participants choose 2032 or 1547. According to them, it is the name of the research project 2000Lakes that misleads them, makes them think there should be at least more than one thousand of recorded lakes.

We evaluate the game by the last question, where participants are asked to rate how much the game makes them want to learn more about mountain lakes. As shown in Figure 9, the answers span a wide range form 1 to 7. There are people who like the game and think it is a great motivator for them to learn more about Swiss mountain lakes. But there are also people who do not enjoy playing the game and do not agree that it motivates them to learn at all. Most of the participants hold a moderate opinion towards this game, rating 3 to 5 for the answer. This game does not meet everyone's satisfaction. But these divergent opinions are very interesting, and we hope to improve this little game from participants' feedback.

Some positive feedback include:

" I like the the small game the most, it is very interactive and can evaluate the outcome of the learning. "

" The game is really cool, because you can learn a lot of things, it's really useful, a very good idea. "

The goal of this educational game is recognized by some of the participants. Learning through game by revealing correct answers seems to be a good way to make players learn.

There are also some suggestions on how to improve the game. For example, one comment saying:

" Maybe for the last game, eliminate those lakes do not have known parameters. Then it will be clearer. "

It is a good suggestion as missing data indeed restricts the questions that the chat bot can ask. If the bot picks one lake with many missing parameters, then the correct answer may be not so informative and the player can only guess the lake purely by its location. In this case, the player cannot really learn new things about the lake. But since there are very limited lakes with complete information, eliminating incomplete ones may leave the game with few options. We therefore hope to add more information to complete current dataset in the future.


Fig. 9. Distribution of answers for Q10: How much does the game makes you want to learn more about mountain lakes in Switzerland?

## 4.2 Aesthetic Design Evaluation

After the tasks, participants are asked to evaluate the aesthetic design of the platform. They rate each of the items in the scale from 1, strongly disagree, to 7, strongly agree. Almost all of the participants agree that everything goes together on this website and the layout is pleasantly varied. And the answers for these two items are all 4 or above, with over a half of the participants showing pretty strong agreement (>=6).

The results for "the color composition is attractive" and "the layout appears to be professionally designed" are not so consistent, as shown in Figure 10. Even though most of the participants agree with these two statements, the color composition is not very attractive for all of them and there are two participants disagree that the layout is professionally designed.

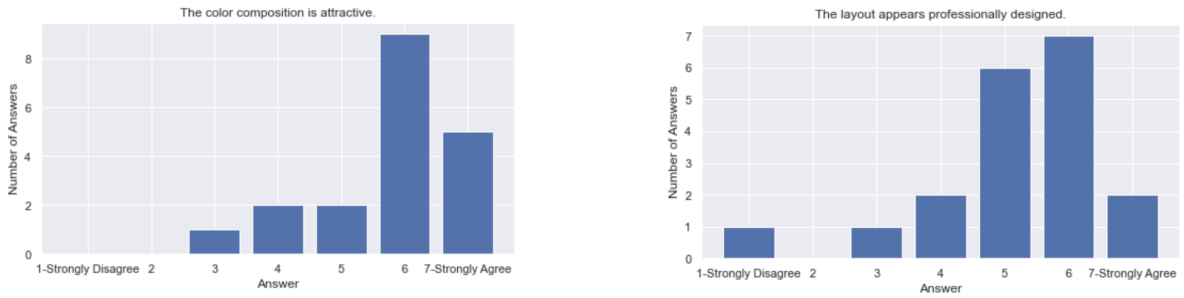



Fig. 10. Left: answer distribution for "the color composition is attractive". Right: answer distribution for "the layout appears to be professionally designed".

Some feedback on color usage and layout include:

“The colors were hard to see sometimes. (clicking on Canton Vaud, I can barely distinguish where the lakes are, I don’t know if it is my settings but the color of the blue circles is very light.)”

“The layout before the game. Some words overlap with the pictures.”

The problems may be caused due to different settings on different devices. The layout issue for example, may be caused by different screen resolutions. Even though each of the web elements is sized in terms of percentage, which means they are scalable on different size of the window, the font size is fixed, and caused the overlap issue when the container becomes too small. It is our goal to make this platform resilient to all circumstances. Therefore, some modifications in terms of layout design need to be done. Also, a clearer color should be chosen to plot the lakes on maps. As there are different map layers, where the background color varies from light to dark, it is important to apply a color that is obvious in all cases or apply different colors for different background.

## 4.3 User Experience Evaluation

Based on the responses from the survey, the user experience this platform provides is overall good. Most of the participants think there are enough supportive information for them to finish the tasks, and it is efficient to find information. Also, most of them think the platform is interesting and exciting to play with. To make it more specific, some participants answer the optional open questions in the end and saying that they favor the interactive map and the section where detailed lake information is displayed. Showing photos seems to be a good way to present the lake as some participants comment that real photos of lakes are interesting, and they enjoy seeing all the info and pictures.

However, there are some participants find this platform not so inventive and innovative. And even though most of them agree that this platform is clear, there are still few participants voting for confusing. The item,

"evaluate based on easy versus complicated" gets more negative responses than the other items and the results are shown in Figure 11. To improve the platform, it is desired to make the operations easier for all kinds of users.

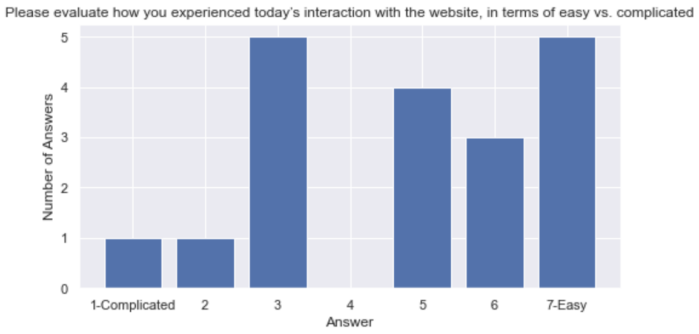



Fig. 11. Answer Distribution for "evaluate how you experienced today's interaction with the website, in terms of easy vs. complicated".

Some participants also give suggestions on how to improve the platform:

"Too much text, it would be nice to have summary or key words for each paragraph. Scrolling up and down is inefficient, it could be more comfortable to have detail information right next to the map."

"On the map you could put a bar where you can adjust the zoom level on the map, because going from one canton to another is sometimes complicated if you have set the zoom."

Even though the paragraphs have been simplified and made more comprehensible, there are still long texts for user to read. Highlight keywords by adding bold style to them is a possible approach to shorten the reading time. Also, we may need to explore a new way to make the interaction with the map more efficient and easy. Currently, if the user clicks on one lake from the map, the page will relocate the user to the lake section automatically, if the user wants to view another lake, he or she has to scroll back to the map section and continue operations. To zoom in to another canton, the user needs to click on target canton directly or first zoom out from current canton and then click on the target canton to zoom in. As some of the users identify these operations to be complicated and inefficient, we may need to find an approach to eliminate redundant steps.

Other feedback include:

"Scrolling is problematic, also we don't know we have to scroll down first."

"The points representing for lake by coordinate is a little bit small to catch."

To address these issues, the scrolling function should be made more explicit, for example, a scroll bar or a list of radio buttons can be added on the side to indicate which section the user is on the website. When plotting lakes on the map, the size of the points should be well adjusted so that it will be easier for users to interact with.

## 4.4 Platform Influence Evaluation

We evaluate the outcomes after participants have experienced this platform. All participants agree to different extends that they learn something new about Swiss alpine lakes after the exercise and 8 out of 19 participants strongly agree. Around two thirds of the participants show interest in learning more on this topic and the rest hold a neutral attitude. Most of them agree that they become more aware of how people can affect the lakes and more concerned about the need for environmental protection. However, the percentage drops slightly compared to the previous two outcomes. We may need to add more contents stating the link between lake ecosystem and the environment to impress the citizens.

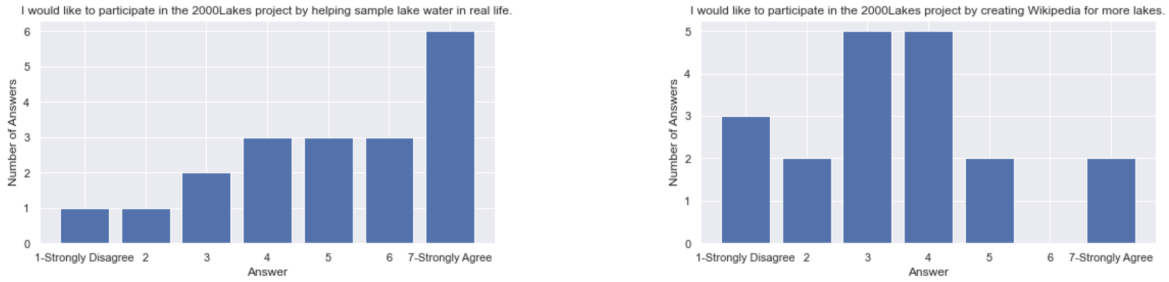



Fig. 12. Left: answer distribution for "I would like to participate in the 2000Lakes project by helping sample lake water in real life.". Right: answer distribution for "I would like to participate in the project by creating Wikipedia for more lakes".

We also evaluate the possible next steps for the participants. Over two thirds of the participants are willing to share this platform to friends or family. However, when asked if they would like to attend sampling activity and if they would like to create Wikipedia for more lakes, their answer span a wide range from 1, strongly disagree, to 7, strongly agree (Figure 12). Even though over half of the persons are interested in participating in sampling lake water, there are some people disagree that they will join. This may be related to their personal interest, as those participants who select 1 to 3 for this statement are those who seldom go hiking. From the graph, we can see that participants are more reluctant to create Wikipedia page compared to sampling water, with most of the them disagree that they will write Wikipedia for more lakes. The reason may be writing Wikipedia page is more task-oriented with no social engagement compared to hiking and sampling water. Another reason may be the exposure of Wikipedia on this platform is not so obvious. For future work, we can explore other interesting ways to invite users to draft Wikipedia page for Swiss mountain lakes.

## 4.5 Comparison with Related Work

We compare the results mainly with the project described in [15], which aims to support volunteers to collect and submit data about invasive species. Map visualization is also an essential part on this platform and it was evaluated in a similar way. Yet, the project is of a different sort, and therefore, the tasks for participants to finish are different. In [15], volunteers were asked to perform a series of more complicated tasks on the map such as editing map layer, creating species location map etc. Volunteers encountered more problems than the researchers expected and the completion rates for map tasks varies from 25 to 75%. Participants expressed the needs of having simplified functions, more comprehensible icons to help with the operations. In our project, the interactive functions on map are easier. From the results, the completion time for each task is reasonable and the accuracy of all questions are pretty high (Table 2 & Table 3). Indeed, simpler features will allow participants to have better performance. Considering not all participants agreed that our platform is easy enough (Figure 11), we still need to improve the platform towards this direction.

The implemented educational game in this project is not entertainment oriented like the "game-like" artifact implemented in the Citizen Sort project [19] but more similar to a "tool-like" game. The format of player answering multiple-choice questions is similar to the one implemented in the Galaxy Zoo [21] but gaming features were added to make it more enjoyable to play. We underlined the three characters of a successful educational game proposed by Malone [12]. Challenge of this game is ensured by asking player to guess the randomly selected lake. Fantasy is realized by having the player interacting with a chatbot. Even though it is a single-player game, the player will be placed in a social situation through having conversation with the bot. We hope to raise players' curiosity by revealing the lake's attributes one after another. And the final confetti effect is hoped to give them

self-satisfaction. From the results we got, this game seems to be the most controversial feature on the platform (Figure 9). Based on the feedback we received, further improvements are needed.

## 5 CONCLUSION

This project aims to build an interactive platform for 2000Lakes, which is a research project on analyzing the bacteria communities inside Swiss alpine lakes in response to climate change with citizen science purpose. It is hoped to spread general knowledge about mountain lakes in Switzerland, keep citizens informed of the progress and results of 2000Lakes, invite them to join the project by taking various actions, and ultimately build a sustainable society with help from the community.

A web-based platform has been implemented for these purposes. As 2000Lakes itself is related to several scientific concepts, which the general public may be unfamiliar with, it is critical to consider how those scientific data can be depicted to laypeople. This project tries to address this issue in terms of data, visualization, and evaluation. Besides the scientific data from 2000Lakes, the main data used are from Wikipedia, which includes many attributes that do not need additional explanation. Visual data and geocoordinates data of the lakes are also included for better presentation. The platform is designed to be visually appealing and aims to attract and keep the users by various interactive functions. As platform implementation usually goes through a development cycle, valid feedback are sought for improvement. A survey is therefore conducted to evaluate the platform and collect feedback.

The results show that participants can grasp the knowledge well from the platform, and generally speaking, the website provides a good user experience. The outcomes of the influence evaluation further demonstrate that this platform achieves the goals of citizen science. However, participants identify some issues with the current platform. Possible improvements are discussed and should be implemented in the future work.

## ACKNOWLEDGMENTS

This work would not have been possible without the supervision of Prof. Daniel Gatica-Perez, who gave many useful suggestions throughout the whole research process. Special thanks to Dr. Anna Carratalà, for explaining the parameters about microbial abundance and for verifying the scientific concepts displayed on the platform. Thanks also go to members of the social computing group and researchers in 2000Lakes project, they all gave valuable feedback and advice for this platform development during meetings.

## REFERENCES


[1] 2021. Conductivity. https://www.enr.gov.nt.ca/sites/enr/files/conductivity.pdf

[2] 2021. Dissolved Oxygen (DO). https://www.enr.gov.nt.ca/sites/enr/files/dissolved_oxygen.pdf

[3] 2021. pH. https://www.enr.gov.nt.ca/sites/enr/files/ph.pdf

[4] 2021. Water temperature - gov. https://www.enr.gov.nt.ca/sites/enr/files/water_temperature.pdf

[5] Rita Adrian, Catherine M O'Reilly, Horacio Zagarese, Stephen B Baines, Dag O Hessen, Wendel Keller, David M Livingstone, Ruben Sommaruga, Dietmar Straile, Ellen Van Donk, et al. 2009. Lakes as sentinels of climate change. *Limnology and oceanography* 54, 6part2 (2009), 2283–2297.

[6] Alexander Anwyl-Irvine, Edwin S Dalmaijer, Nick Hodges, and Jo K Evershed. 2021. Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior research methods* 53, 4 (2021), 1407–1425.

[7] Glen George. 2010. The impact of climate change on European lakes. In *The Impact of Climate Change on European Lakes*. Springer, 1–13.

[8] Yen-Chia Hsu, Jennifer Cross, Paul Dille, Illah Nourbakhsh, Leann Leiter, and Ryan Grode. 2018. Visualization tool for environmental sensing and public health data. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*. 99–104.

[9] Yen-Chia Hsu, Jennifer Cross, Paul Dille, Michael Tasota, Beatrice Dias, Randy Sargent, Ting-Hao Huang, and Illah Nourbakhsh. 2019. Smell Pittsburgh: Community-empowered mobile smell reporting system. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 65–79.

[10] Yen-Chia Hsu, Paul Dille, Jennifer Cross, Beatrice Dias, Randy Sargent, and Illah Nourbakhsh. 2017. Community-empowered air quality monitoring system. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1607–1619.

[11] Edith Law and Luis Von Ahn. 2009. Input-agreement: a new mechanism for collecting data using human computation games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1197–1206.

[12] Thomas W Malone. 1980. What makes things fun to learn? Heuristics for designing instructional computer games. In *Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*. 162–169.

[13] Thomas W Malone and Mark R Lepper. 2021. Making learning fun: A taxonomy of intrinsic motivations for learning. In *Aptitude, learning, and instruction*. Routledge, 223–254.

[14] Morten Moshagen and Meinald Thielsch. 2013. A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology* 32, 12 (2013), 1305–1311.

[15] Greg Newman, Don Zimmerman, Alycia Crall, Melinda Laituri, Jim Graham, and Linda Stapel. 2010. User-friendly web mapping: lessons from a citizen science website. *International Journal of Geographical Information Science* 24, 12 (2010), 1851–1869.

[16] Annu-Maaria Nivala, Stephen Brewster, and L Tiina Sarjakoski. 2017. Usability Evaluation of Web Mapping Sites. In *Landmarks in Mapping*. Routledge, 239–256.

[17] Victoria Palacin, Sarah Gilbert, Shane Orchard, Angela Eaton, Maria Angela Ferrario, and Ari Happonen. 2020. Drivers of participation in digital citizen science: Case Studies on Järviwiki and safecast. *Citizen Science: Theory and Practice* 5, 1 (2020).

[18] Nathan Prestopnik and Kevin Crowston. 2012. Purposeful gaming & socio-computational systems: a citizen science design case. In *Proceedings of the 17th ACM international conference on supporting group work*. 75–84.

[19] Nathan R Prestopnik and Kevin Crowston. 2011. Gaming for (citizen) science: exploring motivation and data quality in the context of crowdsourced science through the design and evaluation of a social-computational system. In *2011 IEEE Seventh International Conference on e-Science Workshops*. IEEE, 28–33.

[20] R Psenner. 2003. Alpine lakes: Extreme ecosystems under the pressure of global change. *EAWAG news* 55 (2003), 12–14.

[21] M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. 2009. Galaxy zoo: Exploring the motivations of citizen science volunteers. *arXiv preprint arXiv:0909.2925* (2009).

[22] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence, 4 (6), 103-108.* (2017).

[23] Wikipedia contributors. 2022. IPCC Sixth Assessment Report — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=IPCC_Sixth_Assessment_Report&oldid=1093620641 [Online; accessed 23-June-2022].

[24] Craig E Williamson, Jasmine E Saros, Warwick F Vincent, and John P Smol. 2009. Lakes and reservoirs as sentinels, integrators, and regulators of climate change. *Limnology and Oceanography* 54, 6part2 (2009), 2273–2282.

## A DISTRIBUTION OF RECORDED LAKES


## B QUESTION POOL FOR THE GAME

All the questions in the question pool are listed here. From these questions, the chat bot will randomly pick four and ask the user. The player needs to select an answer to finish the sentence or they need to decide if the statement is true or false.

- The elevation of this lake is in the range (in meter) ...
- The surface area of this lake is ...
- This lake has its own wiki page.
- The measured pH value of this lake ...
- The measured conductivity of this lake ...
- The measured temperature of this lake ...
- The measured dissolved oxygen level of this lake ...
- The max length of this lake is ...
- The max width of this lake is ...
- The max depth of this lake is ...
- The water volume of this lake is ...

## C EVALUATION SURVEY

# 2000Lakes Platform Evaluation

2000LAKES is a seed project supported by the CLIMACT Research Center, whose goal is the study of climate change and its impact, aiming for a sustainable future. The project is a collaboration involving Idiap, EPFL, and UNIL.

In Switzerland, there are more than 1500 lakes located above 2000 meters of altitude. In order to understand the ecological impacts of climate change on bacteria communities in high mountain lakes, researchers analyze their chemistry and biology.

Within the scope of this project, this interactive platform has been created to promote awareness about Swiss mountain lakes and to conserve these ecosystems by joining forces with local citizens.

Please open **https://linyhazel.github.io/2000Lakes_web/** in your browser **with full screen** to ensure your experience. You are invited to realize a series of interactive tasks via your browser related to finding information about Swiss lakes, and answer a few questions to evaluate the interactive platform. Your participation will help improve its design and bring insights to the research team.

The data to be collected only includes your **anonymous timestamped answers to the survey**. No other data is collected.

The participation in this experiment is voluntary. As a participant, you have the right to withdraw from the project at any time, without any negative consequences for you.

All data processed in this research project will be collected and stored during five years securely and anonymously at Idiap in accordance with the Swiss Federal Act of Data Protection (FADP 235.1).

It takes about 20 minutes to complete this survey.

There are 28 questions in this survey.

## Consent Form

I understand the purposes of the experiment as described in the previous Information Sheet and confirm that it has been explained to my satisfaction.

I accept to participate voluntarily.

I understand that I can decide at any time to no longer participate in the research project without giving reasons and without any negative consequences for me. In this case, it is enough to communicate my decision to the researchers.

I have been informed that the data to be collected includes my anonymous timestamped answers to the survey, and that all data processed in the research project will be collected and saved securely and anonymously at Idiap for five years, in accordance with the Swiss Federal Act of Data Protection (FADP 235.1).

I accept that the Idiap principal investigator (Prof. Daniel Gatica-Perez) and his research team have access to the original data.

In case of doubt, or in case of discomfort related to my participation in this study, I will contact directly the researcher conducting this study (Yuanhui Lin, email: yuanhui.lin@epfl.ch).

*By checking this box, you confirm having read, understood and agree with the Information sheet and the Consent form. If you do not agree to participate, you can leave this page.

Check all that apply

☐ I confirm

# General Information

Please provide some basic information about yourself.

*How concerned are you about the environment?

| | 1-Not at all | 2 | 3 | 4 | 5 | 6 | 7-Very much |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*How often do you go hiking in Switzerland each year between April and September?

| | Never | 1-2 times | 3-5 times | 6-8 times | 9-11 times | 12-14 times | 15 or more times |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*How much do you know about Swiss mountain lakes?

| | 1-Very little | 2 | 3 | 4 | 5 | 6 | 7-A lot |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Can you please specify your age group?

Choose one of the following answers

- ○ 18-24
- ○ 25-34
- ○ 35-44
- ○ 45-54
- ○ 55-64
- ○ 65 and over

*How do you identify your gender?

Choose one of the following answers

- ○ Woman
- ○ Non-binary
- ○ Man
- ○ Prefer to self-describe: _____

## Tasks

Website link: **https://linyhazel.github.io/2000Lakes_web/**

Please follow the instruction and find the right answer to the question.

*__Instruction__: Load the page; observe the first section of the platform.

**Question 1**: How many lakes are recorded in total?

Choose one of the following answers

- ○ 1547
- ○ 2032
- ○ 237

*__Instruction__: Read the fourth slide of the measured parameters.

**Question 2**: For the conductivity parameter, is the higher the better?

✔ Yes ⊘ No

*__Instruction__: Read the fifth slide of the measured parameters.

__Question 3__: Which range of pH value is the most suitable for all kinds of fish?

Choose one of the following answers

- ○ 0-4
- ○ 4-6.5
- ○ 6.5-8.5
- ○ 8.5-11.5
- ○ 11.5-14

*__Instruction__: Scroll down to the next map section of the platform; observe the map by moving your mouse around it.

__Question 4__: Which Canton has the most lakes?

Choose one of the following answers

- ○ Bern
- ○ Valais
- ○ Fribourg
- ○ Graubünden

*__Instruction__: Go to the map section; click on Canton Valais on the map; find the point representing Lac de Salanfe and click on it. You may need to check the information from the first section about dissolved oxygen.

__Question 5__: Lac de Salanfe in Canton Valais is in a very bad condition for the organisms inside in terms of dissolved oxygen. Is this statement true?

✔ Yes ⊘ No

*__Instruction__: Select "Lakes by surface area" from the top left dropdown box; click on Canton Graubünden; observe the largest lake.

__Question 6__: The largest mountain lake in Canton Graubünden is ...

Choose one of the following answers

- ○ Zervreilasee
- ○ Lago di Lei
- ○ Lago Bianco
- ○ Lago di Livigno

30 •

*Instruction: Select "Lakes by elevation" from the top left dropdown box; click on Canton Vaud; observe the highest lake. (Optional: check the google map box on the top right.)

**Question 7**: The highest mountain lake in Canton Vaud is ...

Choose one of the following answers

- ○ Lac des Chavonnes
- ○ Lac de l'Hongrin
- ○ Lac Lioson
- ○ Lac Brenet

*Instruction: Click on Canton Glarus on the map; find Limmernsee and click on it to get more information.

**Question 8**: The water volume for Limmernsee in Canton Glarus is ...

Choose one of the following answers

- ○ 243 million cubic meters
- ○ 93 million cubic meters
- ○ 25 million cubic meters

*Instruction: Here we will compare two lakes. Select "Lakes by coordinates" from the top left dropdown box. First, click on Canton Valais, find Guggisee, and click on it to get the corresponding temperature information; Second, click on Canton Bern, find Oeschinensee, and click on it to get the corresponding temperature information.

**Question 9**: Guggisee in Canton Vaud has a lower water temperature than Oeschinensee in Canton Bern. Is this statement true?

✔ Yes ⊘ No

*Instruction: Play the simple game in the end for at least one round.

**Question 10**: How much does the game makes you want to learn more about mountain lakes in Switzerland?

| | 1-Not at all | 2 | 3 | 4 | 5 | 6 | 7-Very much |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

# Usability Evaluation

Please answer based on your interaction with this platform.

*Please evaluate how visually appealing the interactive website is.

| | 1-Strongly Disagree | 2 | 3 | 4 | 5 | 6 | 7-Strongly Agree |
|---|---|---|---|---|---|---|---|
| Everything goes together on this website. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| The layout is pleasantly varied. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| The color composition is attractive. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| The layout appears professionally designed. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Please evaluate how you experienced today's interaction with the website, in terms of supportive vs. obstructive.

| | 1-Obstructive | 2 | 3 | 4 | 5 | 6 | 7-Supportive |
|---|---|---|---|---|---|---|---|
| Your answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Please evaluate how you experienced today's interaction with the website, in terms of easy vs. complicated. .

| | 1-Complicated | 2 | 3 | 4 | 5 | 6 | 7-Easy |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Please evaluate how you experienced today's interaction with the website, in terms of efficient vs. inefficient.

| | 1-Inefficient | 2 | 3 | 4 | 5 | 6 | 7-Efficient |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Please evaluate how you experienced today's interaction with the website, in terms of clear vs. confusing.

| | 1-Confusing | 2 | 3 | 4 | 5 | 6 | 7-Clear |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Please evaluate how you experienced today's interaction with the website, in terms of exciting vs. boring.

| | 1-Boring | 2 | 3 | 4 | 5 | 6 | 7-Exciting |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Please evaluate how you experienced today's interaction with the website, in terms of interesting vs. not interesting.

| | 1-Not Interesting | 2 | 3 | 4 | 5 | 6 | 7-Interesting |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Please evaluate how you experienced today's interaction with the website, in terms of inventive vs. conventional.

| | 1-Conventional | 2 | 3 | 4 | 5 | 6 | 7-Inventive |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*Please evaluate how you experienced today's interaction with the website, in terms of innovative vs. usual.

| | 1-Usual | 2 | 3 | 4 | 5 | 6 | 7-Leading edge |
|---|---|---|---|---|---|---|---|
| Your Answer | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

## Questions about Your Possible Future Involvement

*Here we explore possible next steps. Please answer after your interaction with this website.

| | 1-Strongly disagree | 2 | 3 | 4 | 5 | 6 | 7-Strongly agree |
|---|---|---|---|---|---|---|---|
| I learned something new about Swiss mountain lakes in this exercise. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| I would like to learn more about Swiss mountain lakes. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| I would like to participate in the 2000Lakes project by helping sample lake water in real life. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| I would like to participate in the project by creating Wikipedia for more lakes. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| I became more aware of how people can affect the lakes and more concerned about the need for envi-ronmental protection. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| I would like to share the link to this website with my friends and family. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

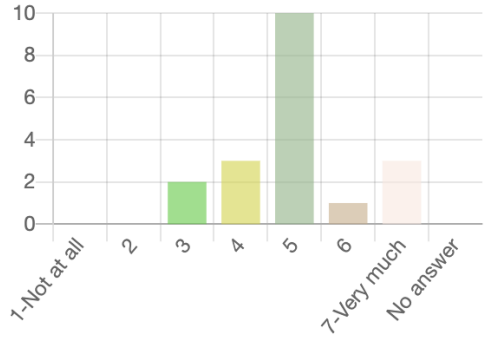
(Optional) Which part of this platform do you like the most?

(Optional) Which part of this platform do you think need improvement? Any advice?

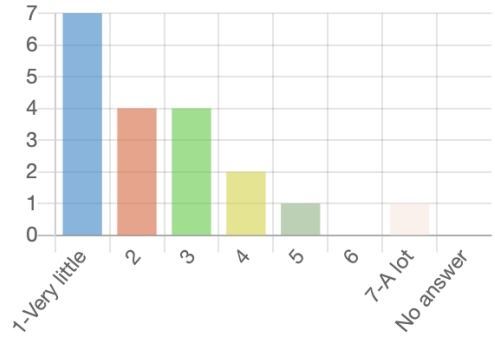
## D SURVEY RESULTS

Results for the survey are shown here. Task results are shown in Table 3 and Figure 9. Open question feedback are discussed in Results section.

(a) How concerned are you about the environment?




(b) How much do you know about Swiss mountain lakes?




(c) How often do you go hiking in Switzerland each year between April and September?

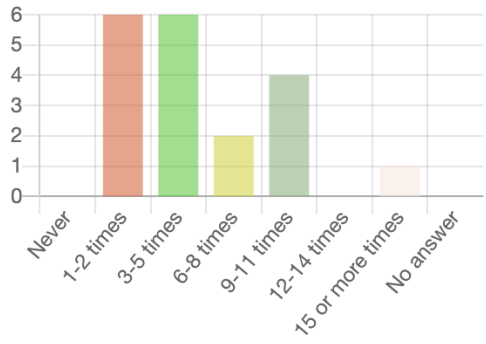
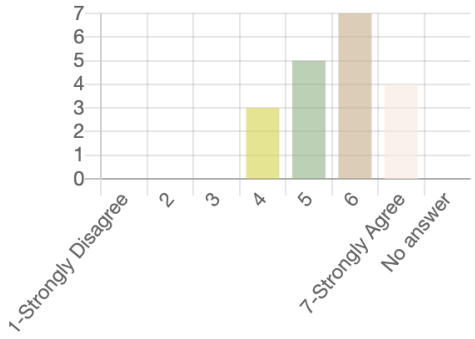
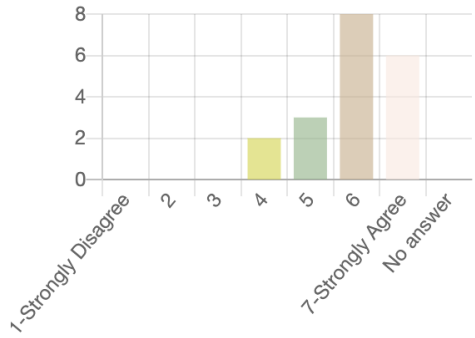



Fig. 13. Results for General Information

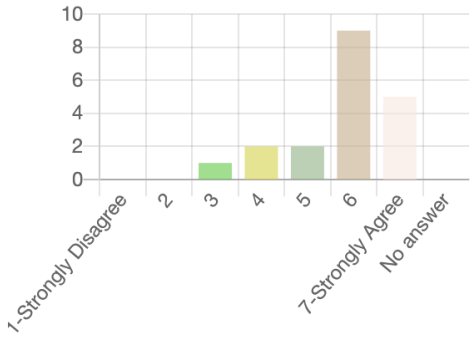
(a) Everything goes together on this website.




(b) The layout is pleasantly varied.




(c) The color composition is attractive.




(d) The layout appears professionally designed.

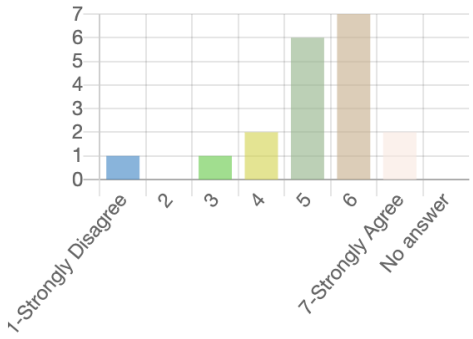
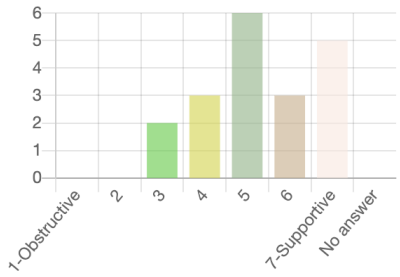
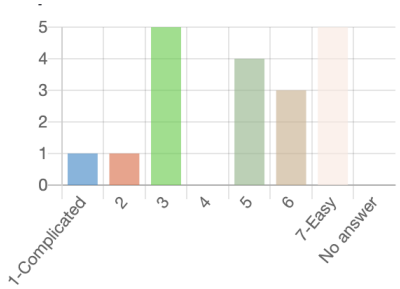



Fig. 14. Results for Visual Design

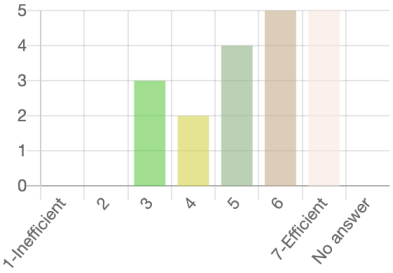
(a) Obstructive vs. Supportive



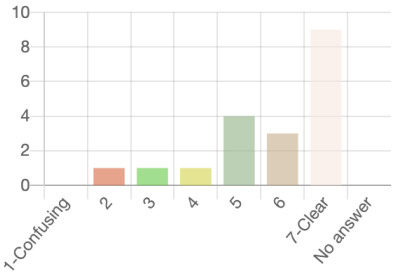

(b) Complicated vs. Easy



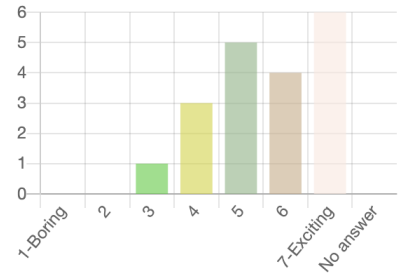

(c) Inefficient vs. Efficient



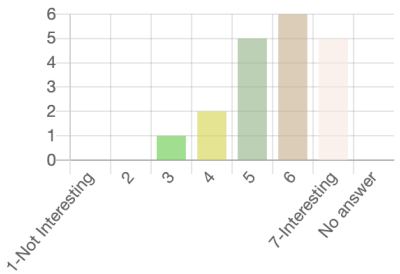

(d) Confusing vs. Clear



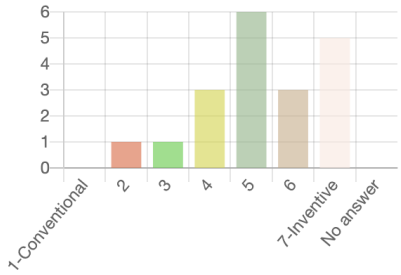

(e) Boring vs. Exciting




(f) Not Interesting vs. Interesting




(g) Conventional vs. Inventive




(h) Usual vs. Leading Edge

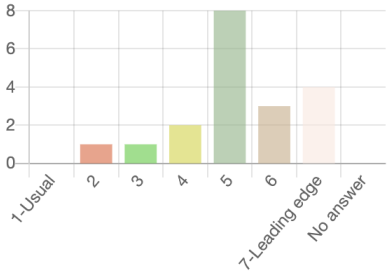
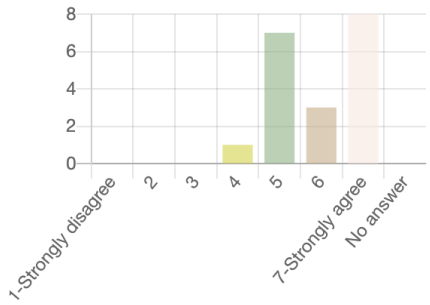
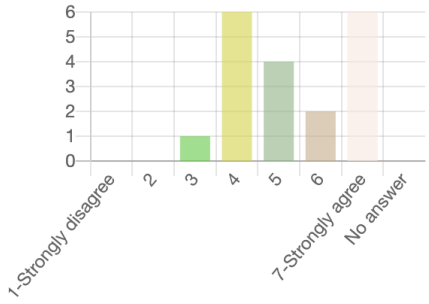



Fig. 15. Results for User Experience

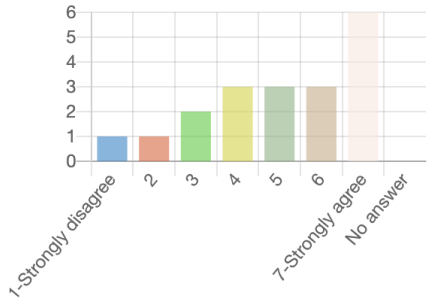
(a) I learned something new about Swiss mountain lakes in this exercise.



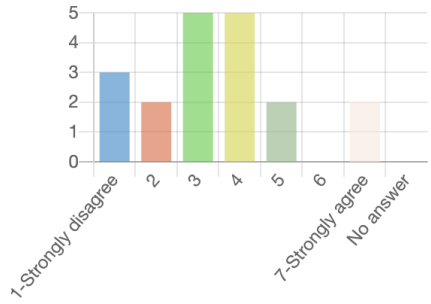

(b) I would like to learn more about Swiss mountain lakes.



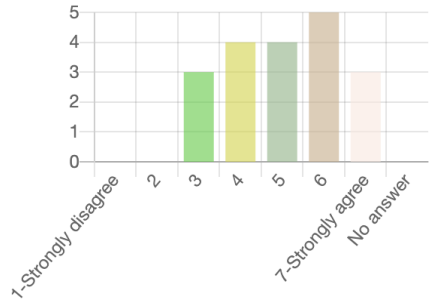

(c) I would like to participate in the 2000Lakes project by helping sample lake water in real life.




(d) I would like to participate in the project by creating Wikipedia for more lakes.




(e) I became more aware of how people can affect the lakes and more concerned about the need for environmental protection.




(f) I would like to share the link to this website with my friends and family.

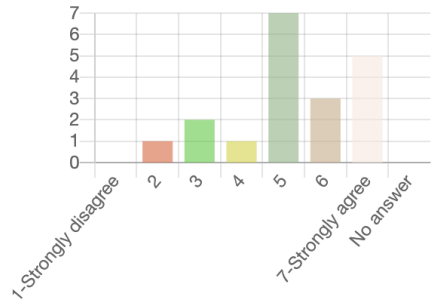



Fig. 16. Results for Influence Evaluation