Characterizing Swiss Alpine Lakes: from Wikipedia to Citizen Science

YUANHUI LIN, EPFL & Idiap Research Institute, Switzerland
DANIEL GATICA-PEREZ, Idiap Research Institute & EPFL, Switzerland

Within the scope of a citizen science project that aims to understand the ecological impact of climate change on bacteria communities in Swiss alpine lakes, we designed and implemented an interactive information platform using data collected from Wikipedia, project-specific data, and other sources. By presenting information about Swiss alpine lakes in an interactive way, the goal of the platform is to raise awareness among the public about the state of Swiss alpine lakes, and ultimately to contribute to the conservation of these ecosystems by engaging citizens. Volunteers were invited to use and assess the platform, by answering questions about alpine lake facts and platform usability. The results show that users can accurately extract factual information from the platform. User feedback was also used to improve the platform functionalities. Finally, an online crowdsourcing activity for lake polygon drawing was conducted to enrich the Swiss alpine lake database with this information. The results show that users can implement this task with high quality.

Additional Key Words and Phrases: Swiss alpine lakes, climate change, citizen science, interactive platform, web-based learning, crowdsourcing

ACM Reference Format:
Yuanhui Lin and Daniel Gatica-Perez. 2023. Characterizing Swiss Alpine Lakes: from Wikipedia to Citizen Science. 1, 1 (August 2023), 24 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Climate change is a global issue of concern. Numerous studies have been conducted to understand its causes and impacts; and strategies have also been proposed to mitigate global warming. From the latest assessment report released by the Intergovernmental Panel on Climate Change (IPCC) [9], the current warmer world is already causing negative effects on natural and human systems. It is important that human societies do not accelerate global warming, and take action to ease the situation.

Lakes play an important role in this issue. They are considered as "sentinels of climate change" [6] because of their sensitivity to environmental factors, especially true for lakes located at high altitudes [37]. Lakes are at the same time witnesses of the past and current climate change and regulators of the future one [45]. To study this important role, previous work has identified several lake variables as indicators of environmental effects [6]. This analysis showed that even the same variables can have different efficacy for different lakes, and as such, different variables should be chosen depending on the research goals.

Many lake studies have been carried out, including the study of single specific lakes [33, 40, 42], lakes within a region [14, 23], and lakes globally [25, 46]. However, alpine lakes in Switzerland remain understudied, despite their significant number. At the same time, Swiss alpine lakes are not often talked about in the local media. As an illustration of this

Authors' addresses: Yuanhui Lin, hazellyuanh@gmail.com, EPFL & Idiap Research Institute, Switzerland; Daniel Gatica-Perez, gatica@idiap.ch, Idiap Research Institute & EPFL, Switzerland.

ACM Journal on Computing and Sustainable Societies, accepted for publication.

situation, an analysis of a database of local French-speaking Swiss newspaper articles in 2015-2022, based on keyword matching and close reading, showed that only 253 news articles were about Swiss alpine lakes, out of a total of over 130,000 news. Discussed topics included activities (132 news) and accidents (52 news) where a specific lake was the location; and the environment (69 news), including issues like energy, climate change, pollution, and biodiversity. In addition, only very few lakes are mentioned by name, compared to the actual number of alpine lakes in Switzerland [5].

The 2000Lakes project [1] aims to better understand the unexplored Swiss alpine lakes and bring the subject to the public. Our partner researchers, with specialized expertise, study the chemistry and biology of these lakes, and are particularly interested in analyzing the abundance of bacteria communities as a response variable to monitor climate change. Other parameters like water temperature, pH value, etc. are also measured. Their research aims to provide necessary knowledge as part of the larger agenda to transition toward a sustainable world.

A second goal of the project is the involvement of citizens through a number of activities, spanning from being informed about the topic, participating in the lake sampling process in the Swiss Alps, and overall becoming more interested in the subject and willing to take action to contribute to environmental protection. As researchers in social computing, we contribute to the citizen science component of the project. In this paper, we designed, implemented, and evaluated a digital platform to assist the 2000Lakes project to achieve its citizen science goal.

Our design aims at building the platform as a digital representation of Swiss alpine lakes. While visualizing project results is a primary goal, various datasets about Swiss Alpine lakes exist in separate sources and disparate formats, and can be integrated, as they can provide a more complete image of the lakes under study. It is also necessary for the scientific data to be presented in forms that laypeople can comprehend. One way to make data engaging for users is by implementing interactive features. The evaluation of the platform is an important aspect of our work. As a citizen science project, how well the public can be involved is an essential criterion for assessing the platform. Having citizens use the platform and comment on its features can offer valuable insights on how to improve it.

The paper presents an analysis of how Swiss alpine lake data can be effectively gathered and visualized for citizen science purposes. We address three specific research questions:

- **RQ1 - Data**: What complementary information can be collected and integrated to facilitate the comprehension of scientific research findings by non-experts?
- **RQ2 - Visualization**: What functionalities can be implemented to make the platform attractive to the public and promote user engagement?
- **RQ3 - Evaluation**: What are effective ways to gather constructive feedback from users, with the aim of assessing and enhancing the platform's functionalities?

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 describes our methodology. Section 4 presents the results. Section 5 discusses the results and their implications. Finally, section 6 provides concluding remarks.

2 RELATED WORK

The concept of citizen science involves people participating in a wide range of collaborative research activities [10, 24, 39]. As the authors of [39] concluded, there are at least three benefits of conducting citizen science. Firstly, people can provide valuable resources to scientists, either data or suggestions on research design and implementation. Secondly, it engages the public and is good for science education and outreach. Thirdly, it democratizes the research process.

[1] https://www.idiap.ch/project/2000lakes/

Citizens can feel as part of the research especially when the research aligns with their own issues of concern. When citizens participate, some ethical issues such as conflict of interest and intellectual property should also be addressed [39].

2.1 Motivations for Citizen Science

It is important to understand the motivations of people for joining research in order to conduct a successful citizen science project. In [11, 28], researchers identified that citizens' willingness to help the environment is a strong factor. Other motivations include their desire to learn new things and to support scientific research. The authors of [34] showed that openness-to-change values, such as the desire to pursue pleasure and novelty, are the main drivers for initial participation. On the other hand, for maintaining sustaining participation, values of self-transcendence, such as the desire to protect everyone's well-being and nature, play a more important role [34].

Many projects use the motivation of learning new concepts and pursuing pleasure. In Galaxy Zoo, volunteers were asked to do morphological classification on galaxy images by answering multiple-choice questions. Researchers then examined the motivations of people taking part in this crowdsourced astronomy project, and found that for this specific project, volunteers mostly participated because of the desire to be entertained, to learn new things, to participate in a community, and to discover something unique [38]. Similarly, in other two projects, Citizen Sort [35, 36] and TagATune [22], games were implemented to amuse users. In Citizen Sort, researchers addressed the issue of attracting volunteers and keeping them involved through games [35, 36]. By implementing a series of artifacts from "tool-like" to "game-like", researchers compared their efficacy and found that the general public indeed showed more interest in the case with game implementation. They also suggested that a good trade-off of entertaining volunteers in scientific research may be adopting the task gamification approach. In TagATune, researchers developed a new game mechanism to let volunteers label audio data [22]. The results showed that the game featuring the new mechanism attracted more participants in comparison to other games that collect audio metadata. As such, effective game design can be considered a critical factor in driving user engagement.

Malone's inspection of the factors that contribute to the enjoyment of educational games is highly informative. As he discussed, there are three essential characteristics that make an instructional game enjoyable, which are challenge, fantasyand curiosity [26]. Later in [27], he extended previous work and introduced more elements like cooperation, competition, and so on. When designing a game, one may take each individual element into careful consideration.

2.2 Map Visualization

Map visualization can be essential when presenting research containing geographical information to citizens. By employing effective visualization methods, the information can be presented in a clear and concise manner, facilitating comprehension. Platforms that embed map implementation for this purpose have been studied in the literature.

In an air quality monitoring system [19] and Environmental Health Channel [17], various data sources are presented to empower citizens in discussing environmental health issues with the authorities. Maps are implemented on the platforms to better indicate the location of issued areas. Interactive elements are overlaid on the map so that users can click and get further information with regard to that region. As another example, Smell Pittsburgh is a mobile application that allows citizens to report pollution odors wherever they are, shows the places where these odors are frequently concentrated, and visualizes odor complaints from citizens in real-time [18]. In this app, the odor reports and the air quality detected by the official sensors are visualized by adding various clickable polygons on top of the map layer. Details will only show after the corresponding polygon is clicked. A general pattern can be found when embedding a map on the platform, extra information is usually added on an overlay above the basic map, and will not show until the request is sent.

In [32], a relatively complete guideline is proposed for map implementation on the web. The authors of [31] also compiled a list of checkpoints to improve map applications, specifically in citizen science projects. Many useful suggestions can be obtained from previous studies, including making the map application simple, adding fun features and help tools, and so on.

In order to enhance the usability of web mapping sites, an appropriate evaluation methodology is crucial. It is possible to evaluate the platform by various means, such as analyzing server logs, conducting online and paper surveys, or interviewing users directly. A hands-on way can be as proposed in [31], which includes drafting a list of map-related tasks, inviting volunteers to finish the tasks, observing their behaviors, and recording the problems encountered.

2.3 Crowdsourcing for Geographical Data

In the context of citizen science, crowdsourcing refers to the practice of having many citizens join data acquisition. By adopting this method, it is easier and less expensive to get a large amount of data, possibly within a shorter time. Moreover, the collected data can cover a wider spatial range and is especially useful to conduct regional research [24].

However, there is an obvious obstacle to crowdsourcing. Data quality is not guaranteed, as data is usually collected by non-professionals, and there can easily be errors or bias [10, 20]. To validate collected data, researchers in [20] suggested comparing citizen-collected data to researcher-collected data, and detecting outliers in this way. In [10], researchers believed that data quality is directly linked to observer quality; therefore reliability can be determined by different profiles of participants. They also underlined the importance of training citizens before letting them collect data by themselves.

OpenStreetMap (OSM) is a large-scale crowdsourced map, constructed by volunteered geographic information. Only registered users can modify the map, in order to trace the information source. Although OSM declares that data quality is ensured because users will rectify incorrect data when using the map, it is not always the case, as researchers found that the users who actually contribute to mapping construction only take up an extremely small percentage [16, 30]. Since OSM does not have systematic, internal data quality assurance procedures [16], it is of interest to investigate the geodata accuracy. Quality assessments on OSM are usually done in a predefined region, collected geodata is compared with authority data in terms of different features [15, 30]. Researchers in [15] highlighted the importance of contributors following a well-defined specification to collect geodata. In the case of OSM, this is possible because there are communities that will hold mapping workshops and there are also plenty of online platforms that support contributors.

3 METHODOLOGY

Our work follows an iterative methodology, wherein valuable datasets from different sources are collected, a base platform is constructed, user feedback is obtained, and subsequent improvements are made to the platform. This section provides a detailed explanation of the methodology adopted at each stage, along with the presentation of the base platform for the first iteration. Additionally, a crowdsourced mapping activity with significance to enrich the data sources for the platform is introduced at the end of this section.

3.1 Data Sources

The data from the 2000Lakes project makes up the first data source of the platform. This scientific data, collected and shared by our project partners, was stored in a structured form, where the name of each sampled lake, its region, altitude, measured pH value, conductivity, temperature, dissolved oxygen, 16S, 18S, g23, and other information were recorded. These measured values can be used to showcase the influence of climate change in Switzerland, offering a fresh perspective on individual lakes. Using this data solely, however, has limitations. To begin with, the data at the time of platform design was limited in terms of quantity, as only 24 lakes had been analyzed with respect to all the above parameters. More generally, it was not realistic to expect researchers to be able to provide complete data in a short time, as sampling and analysis are time-consuming, and also because one of the potential purposes of this platform was to attract people to help with future sampling. Therefore, a database of more Swiss mountain lakes was needed. Secondly, the data variables from the 2000Lakes project are generated from a scientific perspective, which includes concepts that may be unfamiliar to the general public. The terminologies should therefore be well explained on the platform. At the same time, presenting only such scientific data may overwhelm users and make them lose interest. Therefore, additional information that is closer to people's common experience should be added. We addressed **RQ1** by considering various sources of data.

Public, collectively created resources like Wikipedia can be a valuable starting point for citizen science applications [8, 12]. In our case, the Wikipedia page containing a list of mountain lakes of Switzerland [44] is a suitable source of complementary data. More specifically, there are 217 lakes in that list, containing lakes from 15 different cantons. For each lake in the list, the page includes the name, canton, elevation, and surface area. Furthermore, a total of 168 of the lakes have a link to another Wikipedia page, which is about either the glacier or village the lake is located at, or the lake itself. From the lake's own page, some attributes listed in a table were extracted, such as the maximum width, length, depth, and water volume, and were integrated into our database. We also extracted the paragraphs describing the lakes and applied natural language processing (NLP) techniques to them. From an initial assessment, however, named-entity recognition and text summarization did not provide any additional information, so we decided not to use the NLP results. Advanced NLP algorithms were not needed at this initial stage, however, they could be included in the future if they produce additional benefits.

All the data were scraped from Wikipedia using Python with libraries Requests and BeautifulSoup. Out of the 217 lakes, 4 had also been sampled by our research partners and therefore had research results, resulting in 237 lakes in total. The geo-coordinate of each lake was added using the Google Geocoding API, and the accuracy of the coordinates was manually verified.

To obtain more visual information, geo-polygon data of each canton and of each lake was sought. Geojson files of all cantons and lakes were downloaded from Swisstopo (the Swiss Federal Office of Topography) [43], where all the Swiss geodata were made available to the public. 163 out of 237 lakes have such geo-polygon information. Later, an online activity was designed to complete the absent geo-polygon data (see section 3.4.)

Over four thousand images were available from the Wikipedia pages of individual lakes. However, not all lakes have corresponding image information on Wikipedia, and some images dated from a long time ago, which might not present their current appearance. Therefore, other sources of images were needed. We studied other platforms including Google Maps, where photos are uploaded by users. Because a large population uses Google Maps, it has a more comprehensive collection of photos, meaning that more lakes can be found with photos. Consequently, we decided to utilize Google

Canton	AI	BE	FR	GL	GR	TI	JU	NE	NW	OW	SZ	SG	UR	VS	VD
Number of lakes	3	22 (1)	2 (1)	7	64	35	1	1	2	5	4	8	9	69 (22)	7

Table 1. List of Swiss alpine lakes per canton. The number of lakes per canton that have been sampled so far is shown in brackets when the number is not zero. Only lakes from three cantons have been sampled and are part of the current data.

Maps as our resource. 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 were used on the platform. The initial data comes from the 2000Lakes project, and got extended with data from Wikipedia, Swisstopo, and Google API. The final data summary for the recorded number of lakes in each canton is shown in Table 1. However, please note that not all lakes have every attribute, and that the final dataset does not include all alpine lakes in Switzerland. A complete lake dataset is still a goal for future work.

3.2 Base Visualization Platform

To address **RQ2**, it is imperative to use an attractive color composition to the website. Efforts should be directed towards enhancing the clarity of the layout and implementing interactive functions that will effectively engage users.

3.2.1 Aesthetic Design. The aesthetic design of the platform is based entirely on the design of the logo of 2000Lakes project. The font family is chosen to be Montserrat, and the four colors in the logo, one shade of green (#02C39A) and three different shades of blue (#011A38, #90E0EF, #E4F8FB), are applied in different elements throughout the website. Aside from that, the red color (#DA291C) is used, as it is the color of the Swiss flag in the logo. The goal is to achieve a coherent visual appearance between the platform and the project. There are several justifications for this decision. Primarily, the usage of the same logo and design style serves to emphasize its connection with the project. Such coherence reinforces individuals' perceptions of the larger research project. Additionally, it is preferable to reuse the colors from the logo and avoid superfluous colors which could make the platform less visually pleasant. Finally, green and blue colors can be seamlessly associated with the natural environment and are suitable for this particular context.

3.2.2 Function Design. There are certain desired features for the platform. First, a progress bar should be displayed to represent the current status of the project's lake sampling and analysis phases. Showing this to a viewer conveys the message that the project is still at its early stage and volunteer participation in project activities will actually advance its progress further. Ideally, it encourages people to take action and join the research activities.

Second, the analysis results of each lake will be shown on the platform. A section of explanation on these measured parameters is required. Each parameter should be described in detail, including what it is, why it is important for the lake ecosystem, and how human activities can possibly alter it. The influence of human activities needs to be stated explicitly so that users understand their role in protecting lake ecosystems. It is critical that these parameters are not explained in academic terms. Aside from each parameter itself, the actual measured value of each parameter should be explained as well because a pure number carries no meaning for laypeople. It is not necessary to disclose the exact number to the audience, instead the measured value can be divided into several ranges and shown on a scale.

Third, the location of lakes will be shown on a map with different layers. The basic layer is implemented using cantons' geojson file only. Colored polygons will be shown with no other information, and the color indicates the density of lakes in that canton: the darker the color, the higher the density. With this visualization, users get to see roughly the cantonal lake distribution within Switzerland, free from distractions. 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, such as the closest mountain or a nearby town. This can be helpful to locate the lake if users have existing knowledge of that geographic area.

The map can be zoomed in or out, 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 the lakes. Users can choose to visualize lakes based on the coordinates, the elevation, or the surface area depending on their own interest. For example, it is easier to find interesting facts, like the largest mountain lake in a certain canton, if the user chooses to visualize lakes by surface area.

When the computer mouse hovers over a canton or a lake on the map, the basic information of the target should be displayed, eliminating the need for the user to click each time. But if users want further information about a lake, they can click on the lake. Additional data will then be displayed. There should be an entry that directs the users to the lake's Wikipedia page in case they are interested in that particular lake and would like to learn more. For lakes with no Wikipedia page, users will be invited to create a Wikipedia page for that lake based on the information provided. It is hoped that by requesting the help of citizens, the Wikipedia data on Swiss mountain lakes could eventually be completed.

Finally, interesting features should also be implemented in order to attract other citizens. A game similar to the well-known 10-question game is designed for this purpose. 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 our platform, a chatbot will play with the user. The chatbot randomly draws a lake from the dataset. Instead of letting users ask questions, the chatbot generates five multiple-choice questions, lets the user select one answer at a time, and tells the user if the answer was correct or not. All questions will be related to the attributes of a lake, such as its temperature and conductivity value. Users will get to learn about the lake's condition when searching for answers. Since the purpose of the game is to make users learn from the platform and revise what they have learned through browsing, the correct answer is disclosed after each question. It is essential to show players if they won the game at the end of each round, to boost the player's engagement. Ideally, this game will arouse users' interest in learning from the platform, with the hope that the more they become aware of the current state of Swiss lakes, the more likely will be for them to eventually engage in protecting the environment and mitigating climate change.

3.2.3 Implementation and Base Platform. Collected data was preprocessed and constructed in a way that is more suitable for web usage using Python. HTML, CSS, and JavaScript were used to implement the website. The whole implementation also benefits from several external packages and APIs. Bootstrap was used to refine the layout, d3 was used to visualize data and implement various interactive elements, canvas-confetti was used to support the winning effect in the game, and Google Maps API was used to support map functions and display photos of the lakes.

The full set of implemented features is as follows.

When users first load the website, a progress bar is animated to show the status of 2000Lakes, and a note stating the number of lakes sampled versus the number of lakes recorded is displayed as on the top of Fig. 1(a).

Below the progress bar, there is a slider where the scientific parameters are explained. Users can go to different slides to learn about different parameters, including microbial abundance (16S and 18S), temperature, dissolved oxygen, conductivity, and pH value. The explanation of microbial abundance was provided by the domain expert in the

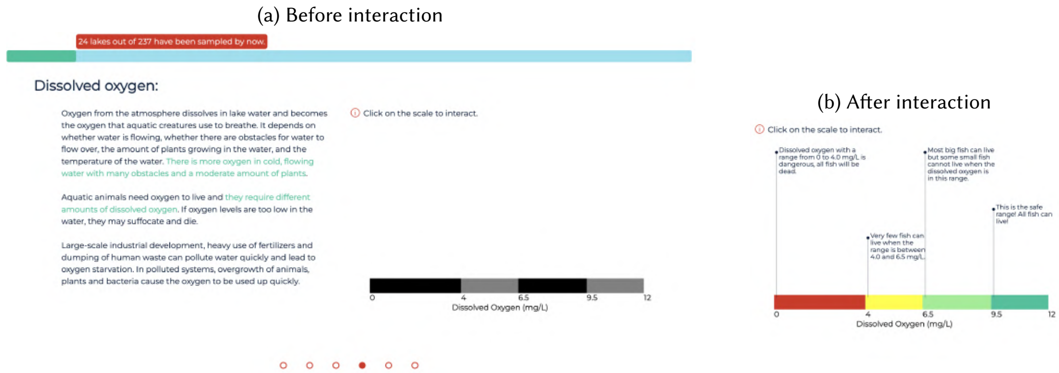


Fig. 1. Progress bar, slider for parameter explanation, and interactive scale.

2000Lakes project. Explanations on water temperature, dissolved oxygen, conductivity, and pH value were drafted based on references [1–4] and the correctness was verified by the expert researcher.

The explanation of measured values is also shown on each slide. Interactive scales were implemented for some parameters in order to make users more engaged in the learning process. For dissolved oxygen, conductivity, and pH value, the division of the scales is based on references [1–3]. By default, the scale is shown as gray and black, with no descriptions as in Fig. 1(a). But if the user clicked on one range, the color changes, and accordingly, a short text describing the meaning of that range appears (Fig. 1, (b)). Different colors apply on different ranges in a way that matches the usual color scheme. Red means dangerous: values in this range cause the death of 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 is in good condition in these cases. It is expected that this implementation could help laypeople better recall 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 temperature scale is divided evenly from 0°C to 20°C, as there are no general 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 from zero to billions, it is not possible to visualize the scale linearly and divide it evenly, therefore these two parameters are displayed in order of magnitude. The exact scales are shown in the lake description so that users can easily remember the meaning (Fig. 3, left).

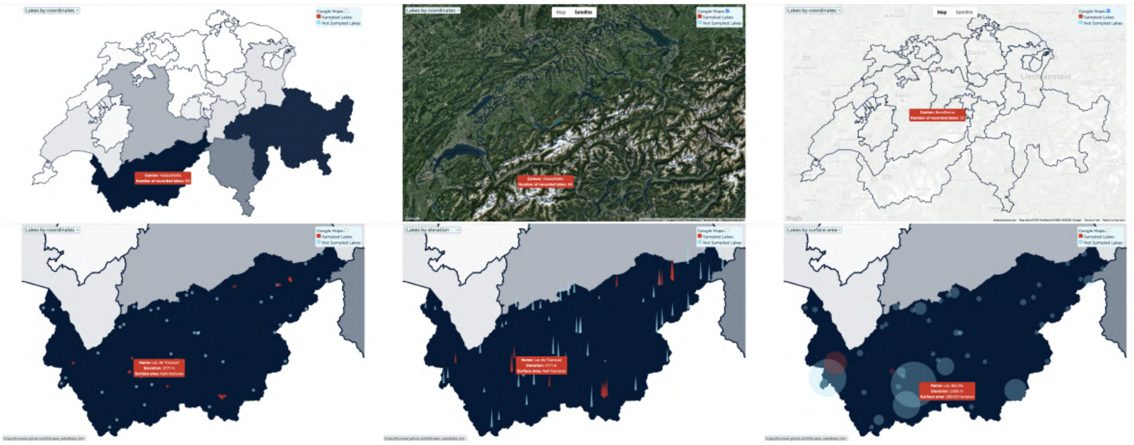


Fig. 2. Upper level: Different map layers. Lower level: Different lake visualizations on the canton level.

Different layers of the map are implemented for different purposes (Fig. 2, upper level). The default map is the one with only colored polygons of all Swiss cantons. Users can also check the Google Maps option. 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 this map. Therefore, it helps users perceive the lakes' elevation. Users can also construct Google Maps with road map tiles, then the main town/city and the main road will be shown. As citizens are usually more familiar with this information, this visualization gives a better idea of where the lake is located.

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

Users can choose to visualize the lakes by coordinates, by elevation, or by surface area from a drop-down box (Fig. 2, lower level). When the mouse hovers over a lake, basic information including its name, elevation, and surface area will be shown in the tooltip. The cursor changes to a zoom-out icon to prompt the user to click to zoom out. Lakes that have been sampled and have not been sampled are colored differently.

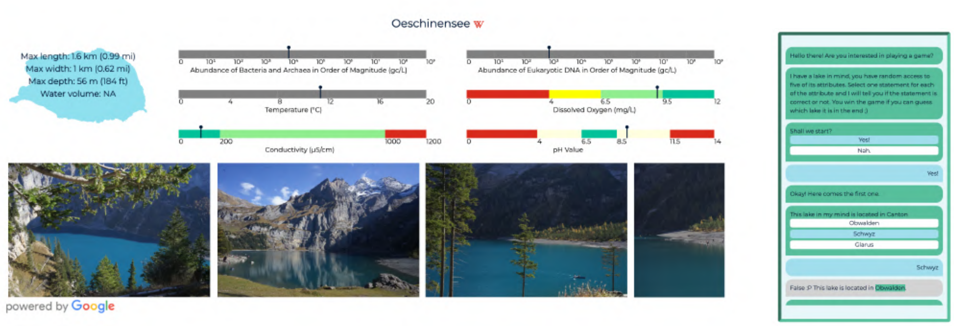


Fig. 3. Detailed lake information and the chatbot game.

When the user clicks on one lake on the map, detailed information about the lake is shown as in Fig. 3, left. The name of the lake is displayed with a Wikipedia icon next to it. Users can click on the icon to check its Wikipedia page, but if the page does not exist, a page written "Hey you found 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 to give the user an overview of the target. Analysis results from the 2000Lakes project are visualized with six scales. Since there are no criteria on how to set the range for 16S, 18S, and temperature, those scales are colored in gray. The rest of the scales are exactly the same as in the explanation section. A small photo gallery of the selected lake, constructed using Google Maps API, is also shown. A maximum of 10 pictures will be shown following Google's restriction. This part gives a more realistic view of the lakes. It is also intended to attract citizens with lake scenes, so that they may be more willing to join future activites, e.g. a sampling campaign in order to enjoy nature, and more likely to take actions to safeguard the environment.

The implemented game is designed like a bot texting with the user in a chat box to make the experience more immersive (Fig. 3, right). When the user gets to this part of the platform for the first time, messages pop up into the chat box. Starting with a greeting, followed by an introduction to the game, the bot then asks if the user wants to play this game or not. Only if the user chooses to play, the bot continues the conversation by giving out questions. The first question is always about the canton where a lake is, and the rest four questions are selected randomly from the pre-defined question pool, where all questions are about the lake's attributes, such as max depth, measured conductivity, etc. Missing data for some lakes results in some questions bearing no useful information. In such cases, the location of the target lake can be very essential for players to guess the right answer. Once the user makes a guess, the bot answers immediately if the guess is right or not, and replies accordingly. For a wrong guess, the reply bubble is set to a different color, and the right answer is highlighted so that the user can distinguish the results more easily and locate the right attribute.

In order to make the dialogue resemble human communication and boost the enjoyment of interacting with the chatbot, the messages sent by the bot are written in a friendly and encouraging way. Several predefined responses were prepared, and the bot randomly picks one to send. The bot is humanized by using a relaxing tone and avoiding fixed replies.

If the user finds out the lake the chatbot selected, a confetti effect is triggered; this is to make the player feel satisfied and pleased through this positive feedback, and to encourage the player to continue learning on the platform and through the game. If the player guesses wrongly, the correct lake is revealed, and the player can match the given attributes to the target lake so as to learn from the game.

After one round of the game is finished, the bot continues to ask if the player wants to play again, and the cycle continues.

Following the initial online survey conducted to gather feedback on the base platform, we made improvements to the features, which are discussed in section 4.2.1.

3.3 Platform Evaluation: Collecting User Feedback

To answer **RQ3**, the platform is evaluated from different perspectives. As an information platform, where users will learn concepts about Swiss alpine lakes, it is of interest to know how well citizens with no previous knowledge are able to grasp such information. In addition, the website is evaluated in terms of user interface (UI) and user experience (UX) design, as the goal is to make the platform easy and pleasing to use. Finally, understanding how influential this website can potentially be in the future should also be evaluated.

An online survey is constructed for evaluation purposes. Participants first read a short description of the project and sign a consent form to continue. Later, participants are asked to provide some general information about themselves, such as how concerned they are about the environment, how often they go hiking in Switzerland, and how much they know about Swiss alpine lakes. Having this information before participants interact with the platform is important, as it can be used to assess the influence and usability of the platform. For example, if a person with a low degree of environmental concern became willing to take action to preserve the environment or join 2000Lakes project after visiting the platform, that would constitute positive feedback. Furthermore, if a person with limited knowledge of Swiss alpine lakes could still find it easy to interact with and find information from the map, that would also represent positive feedback.

The general information part is followed by a task section. In this section, ten questions are asked, whose answers can be found by exploring the content in the platform. 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?

The questions are designed such that participants have to go through each feature on the platform to find the correct answers. This makes sure the participants actually interact with the platform, and acquire knowledge on lakes and the environmental indicators, so that their response on the usability evaluation part is more reliable. Based on the accuracy of the provided answers, it is possible to assess how effective the platform is at conveying knowledge. In addition, the survey also records the time spent on each question; therefore, it is possible to filter out any response that might be filled in randomly (i.e., too fast).

After participants finish all the tasks, a set of questions about the platform's aesthetic design and interaction experience are asked. To evaluate visual aesthetics, participants are asked to which degree they consider 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 the Visual Aesthetics of Websites inventory, proved to be reliable and able to provide a close representation of the full-length version [29]. To evaluate users' subjective impression of their experience with the platform, an optimized version of the user experience questionnaire is adopted, where the 26 items were reduced to 8 [41]. Participants are asked to scale the platform in terms of attributes like supportive vs. obstructive, easy vs. complicated, efficient vs. inefficient, clear vs. confusing, exciting vs. boring, interesting vs. not interesting, inventive vs. conventional, and innovative vs. usual.

Finally, we are interested in how influential this platform could potentially be. We assess this based on the outcomes and possible actions participants may take after visiting the platform. People are asked to scale to which degree they agree that they have learned something new and would like to learn more on this topic. Participants are also asked if they have become more aware of how people can affect alpine lakes, and more concerned about the need for environmental protection. Finally, participants are asked whether they would like to participate in future activities of the project, by helping sample lake water in real life, creating Wikipedia pages for lakes, or sharing the platform with others. These items are selected to capture the effectiveness of the platform in delivering environmental content, moving from awareness to action. The last three items are concrete activities related to the project that can contribute to climate issues.

We implemented the survey using LimeSurvey. The online survey was approved by Idiap's Data and Research Ethics Committee. Participants are asked to finish the survey using their own laptops. Therefore, the testing environment varies from case to case. However, this should not be a problem, as the authors of [7] indicated that different combinations of device and software have only reasonable effects on the accuracy and precision of a web platform's display and response timing.

3.4 Crowdsourcing the Collection of Lake Polygon Data

One important type of geographic data that remains largely unavailable in digital format is the outline of alpine lakes. Such data can provide valuable information to users, as it allows them to visualize the complete shape of the lake, which may not be fully captured in photos. Moreover, the collected outline data holds significant value for research purposes. Regularly gathering this data enables tracking the evolution of a lake over time, thus facilitating the assessment and monitoring of the impact of climate change. Participants were asked to collect lake polygon data using the platform geojson.io. This data collection was also approved by our Institute's ethical committee. The task for the participants was to draw a lake shape following the shore and output the geojson file. Participants are encouraged to preserve the overall shape of the lake when drawing, but the intricate details are deemed as less crucial. To ensure data quality, a detailed specification and a demo video were provided, instructing how to do the work. Each participant was asked to collect data for five lakes, including one with known official data from Swisstopo. In the later evaluation, the official data served as ground truth, so it is possible to measure the precision and recall of the crowdsourced geo-data pairwise with the following equation:

$$precision = \frac{A_{overlap}}{A_{labeled}} \tag{1}$$

$$recall = \frac{A_{overlap}}{A_{authority}} \tag{2}$$

where $A_{overlap}$ is the overlap area of the two geo polygons, representing the correctly identified lake area, $A_{labeled}$ is the area of the polygon that volunteers identified, and $A_{authority}$ is the area of the polygon from Swisstopo. The turfpy python library was used for the actual calculation.

4 RESULTS

The survey results of the two iterations of the platform are presented in this section. All features of the base platform for the first iteration are presented in section 4.1. The second iteration is a modified version based on the survey feedback for the base platform, the results are presented in section 4.2. Finally, the results of the crowdsourcing activity to collect lake polygon data are presented in section 4.3.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Accuracy	84.21%	94.74%	100%	94.74%	89.47%	78.95%	94.74%	100%	68.42%	NA
Time	54.05s	58.29s	36.92s	32.26s	100.51s	61.67s	40.94s	51.23s	123.42s	64.37s

Table 2. Results for each task. Q1-Q9 are objective questions about alpine lake facts. Q10 is a subjective question about the game perceived value.

4.1 First Platform Iteration

The majority of the survey's recipients are university students in Switzerland, who were invited to fill in the survey voluntarily. The first evaluation of the platform is based on 19 responses that were finished.

The survey is expected to be finished in 20 minutes for someone who has never used the platform before. In practice, the time spent on the survey varies for different participants. The average time spent on the survey is 17 min 23 sec, the median time is 14 min 56 sec. All the response times seem to be reasonable, therefore no response was filtered out.

Among the 19 participants, 10 are women and 9 are men; 13 of them are in the age group of 18-24 years old, and 6 of them are in the age group of 25-34 years old. Most of the participants declared that they care about the environment, with 3 of them declaring to be very concerned about this issue. Therefore, the platform may be of interest to these participants. Furthermore, most of them go hiking in Switzerland, but less than 6 times between April and September, and they do not have much previous knowledge about Swiss alpine lakes. That suggests that the correctness of their answers will be significantly based on the information provided on the platform.

4.1.1 Task Results. Participants followed the instruction and found related information to finish the 10 tasks. Except for the final task, where the participants were asked to rate how much the implemented game makes them want to learn more about Swiss alpine lakes, each of the remaining questions has a correct answer. The accuracy of the response is generally very high: 15 out of 19 participants answered the questions with at most 1 mistake, and that includes participants with very little knowledge on this topic.

In order to assess how easy the platform is to use, we examine the accuracy and time spent for each question (Table 2). Regarding the objective questions (Q1-Q9), Q9 is the most time-consuming question, and has the lowest accuracy. The question is "Guggisee in Canton Valais has a lower water temperature than Oeschinensee in Canton Bern. Is this statement true?". To answer this question, the participant needs to find Guggisee in Canton Valais, get its measured temperature, find Oeschinense in Canton Bern, get its temperature, and compare the two values. One possibility of this low accuracy is that participants get impatient finding the target lakes and decide to answer randomly. The second most 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, and check its dissolved oxygen value. And if they are uncertain about the value's meaning, they would need to go back to the section where parameters are explained and check the meaning. These two questions require looking for one specific lake in Canton Valais, where many lakes have been sampled, so it takes more time to find the target. The time spent on the other questions is reasonable, and participants are able to answer each question within one minute. Therefore, we assume that it is mainly the searching part, which is designed for this survey specifically, that 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 what is the most suitable pH value for all kinds of fish. They both instruct the users to read the explanation on the corresponding parameter and engage with the scale. After interacting with the first conductivity scale, users become more familiar with that feature on the platform and spend less time on Q3, accordingly, the accuracy becomes higher.

We also evaluated the game by the last, subjective question Q10, where participants are asked to rate how much the game makes them want to learn more about mountain lakes. As shown in Fig. 4, the answers span a wide range from 1 to 7. There are people who liked the game and think it is a good motivator for them to learn more about Swiss mountain lakes. There are also people who did not enjoy playing the game and did not agree that it motivates them to learn. Most of the participants held a moderate opinion towards this game, rating 3 to 5 for the answer. In other words, the game did not meet everyone's satisfaction, but these divergent opinions are interesting, and we hope to improve the game based on the participants' feedback.

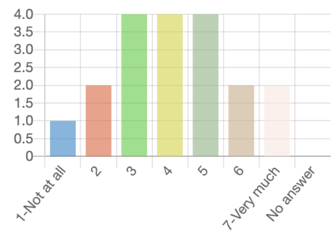


Fig. 4. Distribution of answers for Q10: How much does the game makes you want to learn more about mountain lakes in Switzerland?

Some positive feedback include:

“ I like 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 educational goal of the game was recognized by some of the participants. Learning through the 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 says:

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

Missing data indeed restrict the game. If the bot picks one lake with many missing parameters, the correct answer may be not so informative, and the player can only guess purely based on its location. In this case, the player cannot really learn new things about the lake.

Overall, from the task results, we believe the platform is effective in conveying information to the public.

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

The results for "the color composition is attractive" and "the layout appears to be professionally designed" are varied. Even though most of the participants agreed with these two statements, there are one and two participants, respectively, who disagreed with the statement.

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.”

For future improvement, clearer colors 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

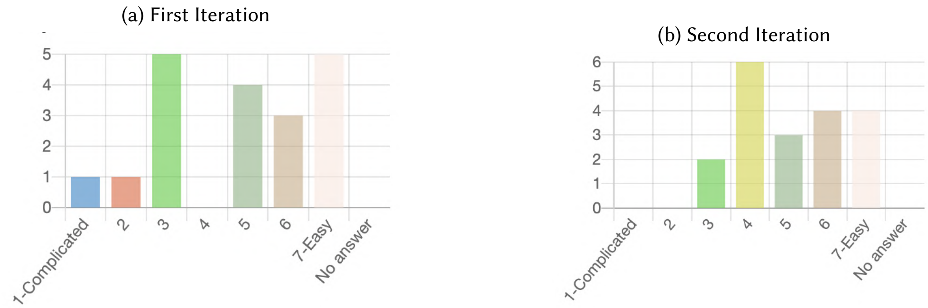


Fig. 5. Result for "evaluate based on easy versus complicated".

apply different colors for different layers. The layout issue, on the other hand, is caused by an inappropriate technical design. Modifications need to be done to make the layout platform more resilient.

4.1.3 User Experience Evaluation. Based on the responses, the user experience provided by the platform is overall good. Most of the participants agreed that there is enough supportive information for them to finish the tasks, and it was efficient to find information. Also, most thought that the platform was interesting and exciting to play with. To make it more specific, some participants said 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 that they enjoy seeing all the info and pictures.

However, some participants found the platform not so inventive and innovative. Furthermore, even though most of them agreed that the platform was clear, there were still two participants rating it as confusing. The item asking to "evaluate based on easy versus complicated" got more negative responses than the other items, and the results are shown in Fig. 5(a). To improve the platform, it is desirable to make the operations easier for all kinds of users.

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

"Too much text, it would be nice to have a summary or keywords for each paragraph. Scrolling up and down is inefficient, it could be more comfortable to have detailed 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."

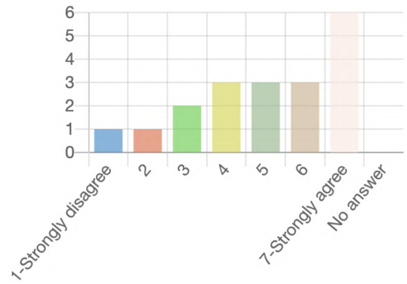
Even though the paragraphs have been simplified and made more comprehensible, there are still long texts for users to read. Highlighting keywords by adding bold style to them is a possible approach to shorten the reading time. Also, we may need to explore a new way to make the interaction with the map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

Other feedback includes:

"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."

(a) Answer distribution for "I would like to participate in the 2000Lakes project by helping sample lake water in real life."



(b) Answer distribution for "I would like to participate in the project by creating Wikipedia for more lakes".

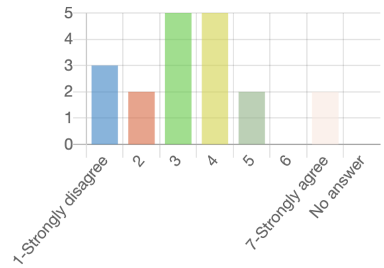


Fig. 6. Result for the first iteration about participants' interest in further activities.

To address these issues, the scrolling function should be made more explicit, for example, by adding a scroll bar or a list of radio buttons on the website. When plotting lakes on the map, the size of the points should be adjusted, so that it is easier for users to interact with them.

4.1.4 Platform Influence Evaluation. All participants agreed to different extent that they learned something new about Swiss alpine lakes after the exercise, and 8 out of 19 participants strongly agreed. Around two-thirds of the participants showed interest in learning more on this topic, and the rest held a neutral attitude. Most of them agreed that they became 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 content stating the link between the lake ecosystem and the environment to be more convincing with citizens.

We also evaluated the possible next steps for the participants. Over two-thirds of the participants seemed 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 spanned a wide range from 1 (strongly disagree) to 7 (strongly agree) (Fig. 6(a)). Even though over half of the persons were interested in participating in sampling lake water, there were some people who disagreed. 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 Fig. 6, we can see that participants were more reluctant to create Wikipedia pages compared to sampling water, with most of them disagreeing that they would write Wikipedia for more lakes. The reason may be that writing a Wikipedia page is more task-oriented and has no social engagement, compared to hiking and sampling water. Another reason may be that the connection between Wikipedia and the platform is not so obvious. For future work, we can explore other ways to invite users to draft Wikipedia pages for Swiss mountain lakes.

In the final stage of the analysis, we intended to understand how participants' demographics influenced their perspectives on the platform. We divided the participants into distinct groups based on gender, age, and their inherent concerns on the environment. Subsequently, we calculated the average scores for each item and present the results in Table 3. Only results with a difference exceeding 0.5 (half a point of the Likert scale) are displayed. We conducted t-tests for each of the 3 conditions being tested (gender, age, and degree of concern for the environment.)

From the table, we observe that women seemed to have higher expectations for UI design as the scores are lower, whereas men seemed to show stricter preferences regarding UX design. Although women appeared to show slightly more empathy and willingness to engage in activities and scored higher for these items than men, the differences are not

	men (9)	women (10)	p-value	18-24 (13)	25-34 (6)	p-value	concerned <5 (5)	concerned >=5 (14)	p-value
Everything goes together	/	/	/	5.85	5.17	0.181	5.00	5.86	0.105
color composition is attractive	6.00	5.44	0.310	/	/	/	/	/	/
layout is pleasantly varied	/	/	/	/	/	/	5.40	6.14	0.146
layout is professional	5.78	4.33	0.023*	5.69	4.00	0.011*	/	/	/
platform is supportive	/	/	/	5.85	4.17	0.007*	/	/	/
platform is easy	/	/	/	5.38	4.50	0.322	/	/	/
platform is efficient	5.00	5.56	0.419	5.54	5.00	0.459	5.00	5.50	0.516
platform is clear	/	/	/	/	/	/	4.80	6.14	0.088
platform is exciting	/	/	/	/	/	/	4.60	5.93	0.039*
platform is interesting	5.11	6.00	0.102	/	/	/	4.80	5.93	0.061
platform is inventive	/	/	/	5.77	4.17	0.020*	/	/	/
platform is innovative	/	/	/	5.46	4.67	0.246	4.60	5.43	0.252
willingness to learn more	/	/	/	5.62	4.67	0.169	4.60	5.57	0.183
willingness to sample water	4.00	5.22	0.205	5.23	3.67	0.120	3.80	5.07	0.238
willingness to create Wikipedia	3.56	3.00	0.468	/	/	/	2.80	3.71	0.328
became more aware	4.56	5.33	0.216	/	/	/	4.00	5.43	0.038*
willingness to share the platform	4.78	5.56	0.273	5.54	4.67	0.245	4.80	5.43	0.432

Table 3. Impact of the demographics of the users. Cases for which the difference between groups is above 0.5 are included in the table. The sample size for each of the groups is indicated in brackets. All p-values below 0.05 are indicated with "*".

statistically significant, as the p-values are all larger than 0.05. The older demographic group appeared to hold higher standards for the platform's overall design, likely due to their increased experience. On the other hand, the younger group demonstrated a relatively greater inclination to participate in various activities, although not significantly. The group with an intrinsic interest in the topic assigned higher scores to the platform and exhibited a stronger desire to participate in all kinds of activities.

It is worth noting that despite some observed differences, the lack of statistical significance in certain cases may be attributed to the small sample size. Beyond our initial work presented here, future evaluations conducted with a larger sample could provide an opportunity to reassess these trends.

	First iteration	Second iteration
Layout professionally designed	5.16	5.42
Complicated (1) vs. Easy (7)	4.79	5.11
Conventional (1) vs. Inventive (7)	5.26	5.26
Usual (1) vs. innovative (7)	5.21	5.16
Willingness to create Wikipedia pages	3.47	3.84

Table 4. Mean value comparison for two iterations.

4.2 Second Platform Iteration

4.2.1 Improvements. Adjustments were made based on the feedback from the first survey results. Despite the minor changes such as adding a scroll bar on the page, adjusting the size of the clickable points, and changing the colors shown on the map, one more interactive feature was added and the layout of the platform was also changed.

The change of layout design intends to address the issue of redundant steps when browsing lakes on the map and when playing the chatbot game. In the first iteration, except for the first explanation section that has horizontal slides, the rest of the sections were joined in a simple vertical order with map visualization followed by the lake information, and the chatbot game was in the last section. When users click on one lake on the map, they will be redirected to the lake information section and need to scroll back to the map section in order to browse other lakes. From the feedback we received, this was not as user-friendly as expected. Moreover, it was hard to compare two lakes because there was only one section that can show information about one lake. Another design that could be improved was the game. Since the game invites the user to interact with the map and find information about a certain lake, it is preferred that the chatbot can be shown somewhere together with the map to make the information retrieval process easier.

In the second iteration, the lake information section and the game section are not shown directly on the main page. When the user clicks on one lake on the map, a new tab will be opened in the web browser. Instead of scrolling, the user needs to switch between the tabs in order to go back to the map section, which is more common in web design. Multiple lake tabs can be opened, so it is easier to compare lakes. Furthermore, the game was integrated into the map section and the chatbot is only shown if the user chooses to open it. In the case of the game being opened, the display window of the map will become smaller and the map will shrink correspondingly to match the window size. The user can then play the game while browsing the map right next to it.

Finally, the goal of the platform is to present information about Swiss mountain lakes, and at the same time invite users to create Wikipedia pages for the lakes based on the information given so to have more complete data online. From the previous survey result, most participants showed a reluctant attitude towards creating Wikipedia pages for lakes (Fig.6(b)). Therefore, we intended to draw users' attention to Wikipedia creation in a more interesting way. We created a virtual lake gashapon machine. When the user clicks on the handler of the machine, a ball will be dropped. Then, the user can click to open the ball, in which there is information about a random lake from the database that has no Wikipedia page. The user is then invited to create a page for that lake. It is hoped that this more innovative way to learn about lakes in Switzerland could trigger users' interest and make them more willing to engage in this activity.

4.2.2 Survey Result. A set of 19 volunteers were invited to answer the exact same survey as for the first iteration; all of them are university students in Switzerland. Since the purpose of the second survey is to verify if the web was indeed improved, items of interest in the survey were analyzed and compared with the results from the first iteration.

(a) Geojson data from Swisstopo (b) Geojson data collected by volunteers

Fig. 7. Visual comparison between official and crowdsourced geo-polygon data.

One major change is the web layout, which intends to simplify some interactions on the platform. Therefore, we are interested in whether users think the layout is professionally designed and if the adjustment indeed makes the platform easier to use. The other change is the gashapon machine, which provides users a way to explore the lake dataset randomly and invites them to create Wikipedia pages. Therefore, we are interested in whether, by adding this new feature, users find the implementation more inventive and more innovative, and if their willingness to create a Wikipedia page is increased. We compare the mean value of the items of interest for the two versions of the platform and the results are shown in Table 4.

An improvement in the mean score can be seen for the items "Layout professionally designed", "Complicated (1) vs. Easy (7)", and "Willingness to create Wikipedia pages". The distributions for "Complicated (1) vs. Easy (7)" are shown in Fig. 5. In the first iteration, participants' answers tended to polarize, with many participants voting for 'not easy'. However, the result improved for the second iteration, with most participants voting for a neutral or positive answer to this item. The score for "Willingness to create Wikipedia pages" also increased. This suggests that the modifications on the platform improved the platform design, the user experience that the platform provides, and user involvement after the experience. However, the mean score for item "Conventional (1) vs. Inventive (7)" stays the same for both iterations, and the mean score for "Usual (1) vs. Innovative (7)" drops a little. While user feedback has acknowledged the presence of the gashapon machine, it did not result in the expected improvement. Therefore, it is worthwhile to explore other methods of interaction for future versions.

4.3 Crowdsourced Lake Polygon Data Quality

A total of 60 lakes were targeted to validate the online crowdsourcing task. 12 volunteers participated, and each of them collected the polygon data of 5 lakes using geojson.io. Even though a detailed specification and a demo video are given, the whole process is unsupervised. Since the data will be shown on the digital platform, the quality of the data should be assessed.

We first randomly sampled from the 60 geo-polygon files volunteers collected and plotted them on geojson.io to see their accuracy manually. The results are representative of the lakes, as the lines match well with the shore shown on the map as shown in Fig. 7. To measure the data quality, precision and recall were calculated. A set of 12 lakes out of the 60 have the official geojson files from Swisstopo, and were used as the ground truth. The collected data reaches an average precision of 95.60% with a standard deviation of 0.061, and an average recall of 97.29% with a standard deviation of 0.016. The crowdsourced data thus shows good quality, and suggests that it is possible to expand the activity and invite more people to join if a larger geo-polygon dataset on Swiss mountain lakes is needed.

5 DISCUSSION

5.1 Simplicity in Platform Design

The simplicity of a platform should be one of the biggest concerns in a citizen science project [31]. Since the potential users of the platform are the general public (who have no obligation to complete any tasks) designing the platform with ease-of-use in mind is critical. On a platform with crowdsourcing purposes, the quality of collected data can be improved; on a platform with informational purposes like ours, users can grasp the information better.

Since map interaction plays an essential role in our work, we are interested in assessing if the design is simple enough for users to interact with the map. We compare the task results from our survey with assessment results for the project described in [31], where map visualization is also an essential part and it was evaluated in a similar way. Note that the project in [31] aims to support volunteers to collect and submit data about invasive species, and therefore, the tasks for participants to finish are different. In [31], volunteers were asked to perform a series of more complicated tasks on the map such as editing the map layer, creating species location map, etc. Volunteers encountered more problems than the researchers expected, and the completion rates for map tasks varied from 25 to 75%. Participants expressed the need of having simplified functions and more comprehensible icons to help with the operations. In our platform, the interactive functions on the map are more straightforward. From the results, the completion time for each task is reasonable and the accuracy of all questions is relatively high (Table 2). Indeed, simpler interaction features allow participants to have better performance.

We also received extensive feedback on various approaches to simplify the platform. Some people mentioned that there was too much text explaining parameters, which made them lose patience. In the improved version, keywords were highlighted in a different color so to assist users' reading. One volunteer also mentioned that the rule of the game was too wordy. Therefore, one way to simplify the platform will be to reduce the text content to a minimum. Additional suggestions were made to reduce redundant scrolling steps while comparing two lakes, as previously described, and to relocate the chatbot game closer to the map interface, given that the game requires the user to search for a lake while playing. Modifying the platform following these suggestions resulted in an improved user experience (Fig. 5).

5.2 Innovative Designs Raises User Interest

Our work sought innovative ways to engage users and raise their interest in this topic. The main efforts are the chatbot game and the lake gashapon machine, whose functions are recognized by the participants.

The implemented game in the platform is not entertainment-oriented, as the "game-like" artifact implemented in the Citizen Sort project [36], but more similar to a "tool-like" game. The format of the player answering multiple-choice questions is similar to the one implemented in the Galaxy Zoo [38], but gaming features were added to make it more enjoyable to play. We underlined the three characteristics of a successful educational game proposed by Malone [26]. The challenge of this game is ensured by asking the player to guess the randomly selected lake. Fantasy is realized by having the player interact with a chatbot. Even though it is a single-player game, the player is placed in a social situation by having a conversation with the bot. We hoped to raise players' curiosity by revealing the lake's attributes one after another. And the final confetti effect was hoped to give them self-satisfaction. Several participants expressed that the game was their preferred feature on the platform, noting that it served as an engaging and informative tool for learning about lakes. In the second iteration, the implementation of the gashapon machine captured the attention of some participants. It then became apparent that these participants were more interested in browsing lakes through the gashapon machine than through the map. If well chosen and implemented, innovative features have the potential to significantly augment user engagement.

5.3 Reliability of Online Crowdsourced Geodata

When it comes to crowdsourcing activities in a citizen science project, researchers in [10] and [15] emphasized the importance of proper training before collecting data and following a detailed protocol while collecting data. In the online crowdsourcing activity we conducted, where participants were asked to collect geo-polygon data using an online tool themselves, a demo video and specifications were made accessible to the participants to show how to do the task correctly. As evidenced by the high quality of the collected data, the provision of clear instructions and strict procedural guidelines were effective. Consequently, it seems feasible to scale up the activity to collect data for more lakes, as long as the critical steps are ensured.

5.4 Challenges of Using Heterogeneous Data

The platform visualizes the research result of the 2000Lakes project. As a bridge connecting the scientific community and the general public, it is important to think about how to present the data in a way that is interesting for laypeople. In addition to interactive features, the use of heterogeneous data presented a viable solution. By extending the research results with relatable data from everyday life, we aim to avoid overwhelming the audience with complex scientific concepts. However, the choice of appropriate data sources and the combination of different sources can be challenging.

As there is no existing dataset of Swiss mountain lakes, our starting point was the Wikipedia page of Swiss mountain lakes. Then, a decision needed to be made on which information should be included on the platform. Our goal was to provide as much information as possible that was directly linked to the topic of interest, while avoiding unnecessary information. Therefore, after a primary NLP process of the paragraphs describing the lakes, we decided not to include the results as they did not provide additional useful information. When confronted with insufficient information from one particular data source, we sought to incorporate additional data from other sources. We made the decision to include more data from multiple sources after a thorough process of trial and comparison, which enabled us to select the most suitable sources for our needs. The combination of different data sources is also tedious work. One way forward is to set up a new dataset that includes every possible attribute of one target. When designing the new dataset, every use case of the data should be taken into consideration, and its scalability should be ensured for the potential addition of new attributes and targets in the future.

5.5 From Citizen Awareness to Action

This study aimed to investigate the effectiveness of a digital platform to engage participants and learn about Swiss alpine lakes in the climate change context, and to motivate them to take action for preserving the environment. The results of the survey indicated an increase in participants' awareness, which shows a possibility of using the platform to gather a community of users who could be motivated to take action. These actions encompass a range of possibilities, including political actions such as writing to representatives, participating in demonstrations, or voting; personal actions such as reducing electricity consumption at home; and professional actions such as engaging in initiatives at work aimed at reducing work-related travel.

However, the transition from awareness to tangible actions still remains a significant challenge. One possible explanation for this challenge is the intention-action gap, which has been widely recognized in the field of behavior change [21]. Despite participants' enhanced knowledge and awareness, converting this newfound understanding into concrete behaviors requires overcoming various barriers and motivational factors. Environmental preservation calls for long-term commitment, lifestyle changes, and collective action, which can be daunting for individuals. As part of future work, we envision that the platform could also integrate historical content to document how the Swiss alpine landscape has changed over time. This could contribute to further raise awareness about the local impact of climate change and motivate people to take specific actions [13].

5.6 Limitations and Future Work

There are improvements that can be made for the lake data this project deals with, for the platform implementation, and for the survey to evaluate the platform.

Firstly, our work encountered an issue with insufficient data on Swiss mountain lakes. An extensive list of lakes is desired not only for this project but also for other potential uses. As we verified the reliability of online crowdsourced geo-polygon data, after a more complete list is acquired, it would be possible to fulfill the missing data of lakes by joining forces with citizens. A more complete dataset will represent benefits for users, regarding learning about lakes in Switzerland.

Secondly, in terms of platform implementation, even though most problems that participants encountered in the first iteration were solved in the second one, the compatibility issue with certain web browsers (Safari) still needs to be addressed. Furthermore, there is the open question of how to make the platform more creative, and how to incorporate additional content regarding climate change and lakes, including individual and social activities around lakes, like hiking and other enjoyable activities. We envision that a potential direction would be through brainstorming sessions with the public to better understand what they expect from the platform. This feedback would help define improvements to our current design.

Thirdly, the evaluations of the two iterations were based on only 19 survey responses, where most of the participants are students. There can be bias in the evaluation of the platform as the participants were relatively few and not diverse enough. In the future, we could expand the sample size and repeat the evaluation with a larger and more diverse audience.

Last but not least, as we envision this platform to be used in the future, it is imperative to not only offer detailed instructions to the public regarding concrete activities, e.g., collecting lake data, but also to provide guidance for calibrating the data, and to leverage the expertise of specialized researchers experienced users. This approach will maximize the platform's potential and overall benefit.

6 CONCLUSION

This paper presented an interactive platform for a research project with a citizen science component towards analyzing microbial life in Swiss alpine lakes as a response to climate change. The platform aims to spread general knowledge about mountain lakes in Switzerland, to keep citizens informed of the progress and results of the research, to invite people to join the initiative by taking various actions, and to contribute to building a sustainable society with help from the community. A web-based platform was implemented for these purposes with a key design requirement: considering how scientific data can be conveyed to laypeople. Our work addressed this issue in terms of data, visualization, and evaluation. Besides project-specific scientific data, the main data in the platform came from Wikipedia, which includes many attributes that do not need additional explanation. Visual data and geo- data of the lakes were also included for better presentation. The platform was designed to be visually appealing and aimed to attract users through various interactive functions. As platform implementation usually goes through a development cycle, feedback was sought for the evaluation of the first iteration through a survey. The same survey was delivered a second time for the improved version, and results were compared in order to validate the modifications. The results showed that participants could effectively grasp facts from the platform, that in general terms the platform provided a good user experience, and that the modifications from the iterative cycle indeed improved the platform. The evaluation of the platform's potential to influence further activities also showed its value to support citizen science goals.

ACKNOWLEDGMENTS

This work was done in the context of the 2000Lakes project, supported by the UNIL-EPFL CLIMACT Starting Grant program. We specially thank Dr. Anna Carratalà (EPFL Environmental Chemistry Lab, Principal Investigator of the 2000Lakes project), for providing the analysis data of the sampled lakes, explaining the parameters about microbial abundance, and verifying the scientific concepts displayed on the platform. We also thank the 2000Lakes project team and the Social Computing group at Idiap, who gave valuable feedback and advice for platform development. We thank Victor Bros (Idiap) for sharing and discussing the Swiss news discussed in the paper. Finally, we thank all the participants in the platform evaluation and geo-data generation phases.

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] 2022. Unpublished analysis of references to Swiss alpine lakes in local Swiss news, conducted by the authors, December 2022.

[6] 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.

[7] 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.

[8] Alexios Brailas, Konstantinos Koskinas, Manolis Dafermos, and Giorgos Alexias. 2015. Wikipedia in Education: Acculturation and learning in virtual communities. *Learning, Culture and Social Interaction* 7 (2015), 59–70. https://doi.org/10.1016/j.lcsi.2015.07.002

[9] Josep G. Canadell, Pedro M.S. Monteiro, Marcos H. Costa, Leticia Cotrim da Cunha, Peter M. Cox, Alexey V. Eliseev, Stephanie Henson, Masao Ishii, Samuel Jaccard, Charles Koven, and et al. 2021. *Global Carbon and other Biogeochemical Cycles and Feedbacks Supplementary Material.* Availablefromhttps://www.ipcc.ch/

[10] Janis L Dickinson, Benjamin Zuckerberg, and David N Bonter. 2010. Citizen science as an ecological research tool: challenges and benefits. *Annual review of ecology, evolution, and systematics* (2010), 149–172.

[11] Margret C Domroese and Elizabeth A Johnson. 2017. Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. *Biological Conservation* 208 (2017), 40–47.

[12] Rosta Farzan and Robert E. Kraut. 2013. Wikipedia Classroom Experiment: Bidirectional Benefits of Students' Engagement in Online Production Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 783–792. https://doi.org/10.1145/2470654.2470765

[13] Christiana Figueres and Tom Rivett-Carnac. 2020. The Future We Choose: Surviving the Climate Crisis (Illustrated Edition).

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

[15] Jean-François Girres and Guillaume Touya. 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS* 14, 4 (2010), 435–459.

[16] Mordechai Haklay and Patrick Weber. 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive computing* 7, 4 (2008), 12–18.

[17] 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.

[18] 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.

[19] 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.

[20] Ashlee Jollymore, Morgan J Haines, Terre Satterfield, and Mark S Johnson. 2017. Citizen science for water quality monitoring: Data implications of citizen perspectives. *Journal of Environmental Management* 200 (2017), 456–467.

[21] Feng-Yang Kuo and Mei-Lien Young. 2008. A study of the intention–action gap in knowledge sharing practices. *Journal of the American Society for Information Science and Technology* 59, 8 (2008), 1224–1237.

[22] 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.

[23] Hye Won Lee, Eun Jung Kim, Seok Soon Park, and Jung Hyun Choi. 2012. Effects of climate change on the thermal structure of lakes in the Asian Monsoon Area. *Climatic change* 112, 3 (2012), 859–880.

[24] Kathryn A Lee, Jonathan R Lee, and Patrick Bell. 2020. A review of Citizen Science within the Earth Sciences: Potential benefits and obstacles. *Proceedings of the Geologists' Association* 131, 6 (2020), 605–617.

[25] Stephen C Maberly, Ruth A O'Donnell, R Iestyn Woolway, Mark EJ Cutler, Mengyi Gong, Ian D Jones, Christopher J Merchant, Claire A Miller, Eirini Politi, E Marian Scott, et al. 2020. Global lake thermal regions shift under climate change. *Nature communications* 11, 1 (2020), 1–9.

[26] 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.

[27] 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.

[28] Phoebe R Maund, Katherine N Irvine, Becki Lawson, Janna Steadman, Kate Risely, Andrew A Cunningham, and Zoe G Davies. 2020. What motivates the masses: Understanding why people contribute to conservation citizen science projects. *Biological Conservation* 246 (2020), 108587.

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

[30] Pascal Neis and Dennis Zielstra. 2014. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet* 6, 1 (2014), 76–106.

[31] 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.

[32] Annu-Maaria Nivala, Stephen Brewster, and Tiina L Sarjakoski. 2008. Usability evaluation of web mapping sites. *The Cartographic Journal* 45, 2 (2008).

[33] Charles G Oviatt. 1997. Lake Bonneville fluctuations and global climate change. *Geology* 25, 2 (1997), 155–158.

[34] 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).

[35] 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.

[36] 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.

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

[38] 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).

[39] David B Resnik, Kevin C Elliott, and Aubrey K Miller. 2015. A framework for addressing ethical issues in citizen science. *Environmental Science & Policy* 54 (2015), 475–481.

[40] Goloka Behari Sahoo, S.Geoffrey Schladow, John E Reuter, Robert Coats, Michael Dettinger, John Riverson, Brent B. Wolfe, and Mariza Costa-Cabral. 2013. The response of Lake Tahoe to climate change. *Climatic Change* 116, 1 (2013), 71–95.

[41] 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).

[42] Scott Stine and Mary Stine. 1990. A record from Lake Cardiel of climate change in southern South America. *Nature* 345, 6277 (1990), 705–708.

[43] Swisstopo. 2022. Free basic geodata. https://www.swisstopo.admin.ch/en/swisstopo/free-geodata.html [Online; accessed Mar. 2022].

[44] Wikipedia. 2022. List of mountain lakes of Switzerland. https://en.wikipedia.org/wiki/List_of_mountain_lakes_of_Switzerland [Online; accessed Mar. 2022].

[45] 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.

[46] R Iestyn Woolway and Christopher J Merchant. 2019. Worldwide alteration of lake mixing regimes in response to climate change. *Nature Geoscience* 12, 4 (2019), 271–276.