

Article

An Integrated Toolbox for the Engagement of Citizens in the Monitoring of Water Ecosystems

Maria Krommyda ^{1,*}, Anastasios Rigos ¹, Spyridon-Nektarios Bolierakis ¹, Theodoros Theodoropoulos ¹, Stefano Tamascelli ², Luca Simeone ²  and Evangelos Sdongos ¹ and Angelos Amditis ¹

¹ Institute of Communication and Computer Systems, National Technical University of Athens, 157 73 Athens, Greece

² XTeam Software Solutions, 45100 Rovigo Veneto, Italy

* Correspondence: maria.krommyda@iccs.gr

Received: 31 March 2020; Accepted: 17 April 2020; Published: 20 April 2020



Abstract: The monitoring of water ecosystems requires consistent and accurate sensor measurements, usually provided from traditional in-situ environmental monitoring systems. Such infrastructure, however, is expensive, hard to maintain and available only in limited areas that had been affected by extreme phenomena and require continuous monitoring. Due to climate change, the monitoring of larger areas and extended water ecosystems is imperative, raising the question of whether this monitoring can be disengaged from the in-situ monitoring systems. Due to climate change and extreme weather phenomena, more citizens are affected by environmental issues and become aware of the need to contribute to their monitoring. As a result, they are willing to offer their time to support the collection of scientific data. Collecting such data from volunteers, with no technical knowledge and while using low-cost equipment such as smart phones and portable sensors, raises the question of data quality and consistency. We present here a novel integrated toolbox that can support the organization of crowd-sourcing activities, ensure the engagement of the participants, the data collection in a consistent way, enforce extensive data quality controls and provide to local authorities and scientists access to the collected information in a uniform way, through widely accepted standards.

Keywords: water ecosystems; citizen science; data quality; integrated platform; smart sensors; innovative sensors

1. Introduction

Environmental monitoring is based on time-series of data collected over long periods of time from expensive and hard to maintain in-situ sensors available only in specific areas. Climate change has prompted the monitoring of extended areas of interest and has raised the question of whether such monitoring can be complemented or replaced by low-cost and easy to use portable sensors and devices. This will allow the collection of needed information with the support of volunteers, enabling the monitoring of aquatic ecosystems and extended areas of interest.

In recent years, scientists have come to understand that widespread access to technological platforms such as smart phones and tablets, with accurate GPS trackers, high resolution cameras and multiple sensors, has opened the way to the general public to participate in scientific research [1,2]. As the collaboration between scientists and citizens in the data collection has increased, the term *citizen science* has gained in popularity.

As a result, people from different social, economic and educational backgrounds are interested in contributing to the collection of data. However, such diverse participation often means that the

volunteers do not have the needed experience and training to collect data in a strict *scientific way*, by following rules and procedures and recording in detail their actions and measurements. *Data collection* is strictly defined as the process of gathering and measuring information on variables of interest, in an established and systematic fashion that enables one to answer stated research questions, test hypotheses and evaluate outcomes [3].

The deviation of the crowd-source data collection from the strict scientific definition has resulted in many scientists expressing skepticism regarding the quality of the collected data from volunteers and questioning the value that these add to their research. The main challenges regarding the use of citizen science [4–7] for scientific research are

Data quality. Critics have raised concerns about data quality, and some studies have shown controversial results. Some studies find that volunteers are inaccurate when tasked with actions that require a great degree of knowledge in a specific field, such as identifying plant species [8]. Other studies, however, indicate that while citizens have a greater learning curve than professionals, they tend to be more focused and dedicated, paying attention to details that professionals may overlook and provide data of the same or higher quality in the long run [9,10].

Communication between volunteers and scientists. Another challenge that materialized from the first efforts of scientists to organize data collections with the help of volunteers was the communication between the two groups [11]. Scientists are used to speaking in scientific terms with which citizens may not be familiar or fully comprehend. This leads to the collection of data in improper ways that do not support the scientific purpose for which they were collected. Volunteers, on the other hand, tend to know better the local topology, the accessibility of the area and the local challenges.

Bias contributions. Concerns have also been raised about the motivations of citizens and their potential biases, which might influence their activities in citizen science projects, especially for cases where the rewards were not direct [12,13]. As many examples have shown, research in general is not devoid from bias [14]. Regardless of their origin, biased contributions should be identified and eliminated from the datasets before the data utilization.

Equipment requirements. Most technological needs for crowd-sourcing data require minimum equipment from the volunteer, a smart phone or tablet, an application and some network connection. There are cases, however, where the scientists want to measure other parameters, such as air temperature and quality, that cannot be provided by equipment already available to the volunteer. In such cases, the scientists are asked to identify sensors, that can be given to the volunteer and facilitate the data collection [15]. While such efforts have proven to be reliable and provide the needed information for the monitoring systems, occasionally, data collected by such sensors may be of low quality or affected by noise, requiring additional quality control measures.

Volunteer engagement and communities of practice. There are many difficulties regarding the engagement of volunteers and the formation of communities of practice [16]. For tasks that require regular and frequent data collections issues about recruitment, commitment and time availability may be a significant barrier. For collection of data to areas that are isolated or not easily accessible, the commute to the area may be disengaging due to the time commitment and the financial burden. In order to mobilize and maintain large citizen science initiatives, it requires the dedication of both time and financial resources. This raised the question of whether investing in citizen science can lower the cost of the data collection process. Recent studies [17] show that the answer to this question is highly dependent not only on the type of data collected but also the type and extent of errors committed by volunteers.

Data discovery and re-usability. Another important challenge regarding data collection through citizen science projects is the data discovery and re-usability. More often than not, data collected in the context of citizen science projects are very specific, measuring data in customized ways that facilitate the collection from people without scientific knowledge and are rarely mapped to generic formats or provided using widely accepted standards. The answer to this problem, comes through the utilization

of the appropriate standards [18], that allow the uniform provision of data between different citizen science projects [19,20].

Contributions. We present here the Scent toolbox, an integrated monitoring system that aims to address all the identified challenges and provide a way for scientists to collect accurate and trustworthy environmental data. A collection of smart collaborative and innovative technologies allow citizens to support scientists and policy makers by collecting environmental measurements. The toolbox has been carefully designed to offer:

Data quality assessment. A data quality control mechanism has been implemented that uses a custom mechanism to identify biased or low quality contributions and remove them from the system.

Communication between scientists and volunteers. The *Campaign Manager* is a web platform dedicated to the unobstructed communication between scientists and volunteers. Scientists can use the tool to define areas and points of interest (POIs) through an interactive map visualization and specify the type of data needed at each point of interest.

Data Assurance. Important measurements for the monitoring of aquatic ecosystems, such as water level and water velocity are collected indirectly. The volunteers are asked to collect images and video, which are processed by dedicated tools that extract the measurements. This eliminates biased measurements, errors due to inexperience or insufficient training and allows for a detailed post-collection scrutiny of the provided contributions.

Integrated portable environmental sensors. For measurements that require the utilization of additional infrastructure, small, low-cost and reliable portable sensors are selected. These provide accurate air temperature and soil moisture measurements with low maintenance needs. The application responsible for the collection of the measurements guides the volunteer and ensures the proper utilization of the sensor while a dedicated quality mechanism ensures the validity of the information.

Volunteer engagement. We have developed a volunteer engagement strategy that is based on gamification. The volunteers can collect points and awards by completing simple tasks, such as taking pictures and video, and contributing to the data collection process.

Data discovery and re-usability. Recognizing the importance of making the collected data available, all the information is modeled, stored and provisioned using the Open Geospatial Consortium (OGC) [21] Standard SensorThings API (Application Programming Interface) [22]. The filtering functionalities of the SensorThings API allow the spatiotemporal discovery of information among the collected measurements and encourage their re-usability.

2. System Architecture

In order to organize data collection activities with the help of volunteers with no technical knowledge, referred to as campaigns, the key for success is to carefully identify the information needed and establish unbiased and reliable processes for its collection in a straightforward and easy way. The established processes should be simple, easy to explain and execute, independent from the knowledge and skills of the volunteer and not dependent on expensive infrastructure. They should also be designed in ways that will ensure the data assurance and provide ample opportunities for data validation and quality control.

The information, in the context of monitoring water ecosystems that has been defined as very important is the monitoring of land cover and land use (LC/LU) changes, the water level and the water velocity as well as the recording of environmental parameters such as air temperature and soil moisture. These needs have been mapped to straightforward, time-effective and meaningful actions that the volunteers can carry out with ease, without requiring special knowledge or skills. The proposed solution for the monitoring of the LC/LU changes is the collection of images of the area of interest and their classification based on a taxonomy to identify depicted elements of interest. For the water level, the volunteer is again asked to capture an image, while for the water velocity the volunteer is asked to collect video. As for the collection of environmental parameters, a small, low-cost but accurate sensor paired with a smartphone or tablet is used.

Aiming to facilitate the users in performing these tasks and address the above-mentioned challenges related to the data collection through volunteers, we have designed an innovative system, which is presented in Figure 1. The system is composed of a suit of components, each one specifically designed to support the tasks assigned to volunteers or to facilitate the data exchange. An innovative flow of information has been implemented to ensure that the needs of the local authorities are communicated to volunteers, while the collected data are forwarded to environmental experts where they are processed and enriched before being presented to the authorities that can now have improved monitoring of the area of interest. The system requirements for the toolbox have been collected through multiple end-user workshops and focus groups, as discussed in [23]. The key components of the system are presented in detail below.

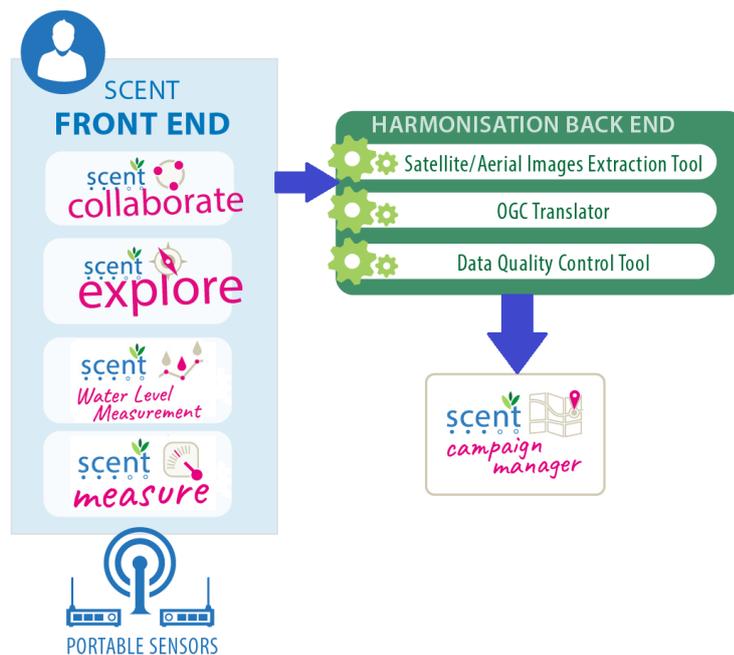


Figure 1. Scent toolbox architecture.

2.1. Lc/Lu Taxonomy

When monitoring an ecosystem, it is very important to have detailed, area specific, information about the land cover and land use. Indicative of the importance of such monitoring are the efforts dedicated by the European Union to the development of a consistent land cover repository. The Copernicus Land Monitoring Service [24] utilizes the CORINE (Coordination of Information on the Environment) Land Cover (CLC) inventory, which was initiated in 1985 and updated in 2000, 2006, 2012 and 2018. It consists of an inventory of land cover in 44 classes. CLC uses a Minimum Mapping Unit (MMU) of 25 hectares (ha) for spatial phenomena and a minimum width of 100 m for linear phenomena. CLC is produced by the majority of countries by visual interpretation of high resolution satellite imagery. In a few countries semi-automatic solutions are applied, using national in-situ data, satellite image processing, GIS integration and generalization. The 44 classes that form the CORINE Taxonomy [25] come from a 3-level hierarchical classification system that respects the detail level of the 25 ha MMU. As an example, for the classification of Water bodies, a level-1 class, there are available two sub-classes, Inland waters and Marine waters. The Inland waters are further divided in Water courses and Water bodies.

In the context of monitoring water ecosystems, however, the CORINE taxonomy does not offer the level of detail required. Specifically, it is of utmost importance to specialize between different types of water courses, wetlands and sparsely vegetated areas. For this reason, the Water bodies class

was further specified to include Dry cross section and River bank. Next the River bank was further specified to include types of land cover at the bank, such as stone, bare soil, concrete and shrubs. All the taxonomy classes used by the Scent toolbox are presented in the Appendix A.

2.2. Campaign Manager

Aiming to alleviate the communication gap between scientists and volunteers, a web-based application, the Campaign Manager, was created. The Campaign Manager allows scientists to communicate their needs regarding data collection activities in a user friendly, simple and functional way. Scientists can use a user-friendly web interface to create campaigns for areas where data on LC/LU, soil conditions and river parameters are needed, while also providing descriptive information about the scope, aim and purpose of the campaigns. In addition, the scientists define for each campaign Points of Interest (PoIs), specific coordinates where data should be collected. Furthermore, the Campaign Manager supports the visualization of citizen generated data, images, videos and sensor measurements, as well as maps of the areas of interest with information regarding LC/LU.

The architecture of the Campaign Manager has been designed according to the classical multi-tier system with a communication paradigm based on open source service-oriented architecture, where each service interacts with the other through a set of messages written in a standard format. The architecture and relevant services are composed of the following layers:

Presentation layer. Mobile and web interfaces, the main gateway for interacting with the application's main services.

Service layer. Services responsible for the communication between parts of the Presentation Layer and also for the communication between Campaign Manager and external API endpoints, further details about these endpoints are provided in Section 5.3, for the acquisition of citizen generated data.

Data access layer. The internal endpoints of the *Campaign Manager*.

Database. A database and collections implemented using MongoDB.

The interfaces of the application, Figure 2, are implemented using Angular 4, a JavaScript-based open-source platform that makes it easy to build applications with the web. Angular combines declarative templates, dependency injection, end to end tooling and integrated best practices to solve development challenges. Moreover, spatial visualizations of the collected data are implemented as map overlays using OpenLayers library along with OpenStreetMaps.

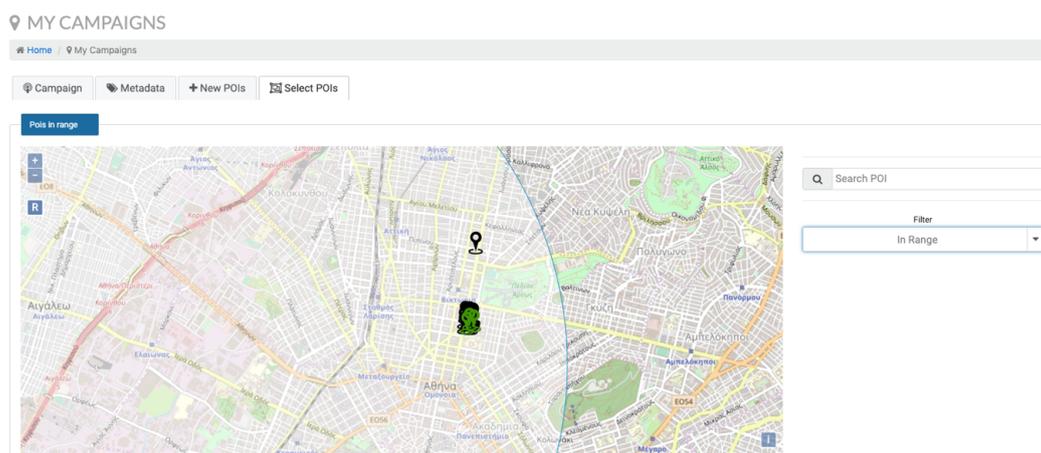


Figure 2. Campaign Manager. PoI visualization.

The server and the RESTful API of the Campaign Manager are implemented using Node.js, an open-source, cross-platform JavaScript run-time environment that executes JavaScript code server-side. Through the API basic user actions, such as authentication and password change, and more complicated ones, such as data CRUD (Create, Read, Update and Delete) actions, are handled. The personalized settings of the user accounts and settings are stored in a database implemented using

MongoDB, a free and open-source cross-platform document-oriented database program, that uses JSON-like documents with schemas.

2.3. Gaming Applications

One of the major challenges in citizen science is keeping the users motivated to contribute over a period of time, in a way that will not require the continuous dedication of time and resources from the scientists. The 1% rule of thumb [26], which is also referred to as the 90-9-1 principle, aims to quantify the imbalance that is associated with the contributions of volunteers to crowd-sourcing efforts. It has been observed that approximately 1% of users of any application or website contribute the vast majority of new content. The percentage that they contribute is calculated in most cases between 90% to 95%. An additional 9% of users are credited with sparse contributions or secondary tasks that may not be as important, depending a lot on the context discussed. An indicative example is *Wikipedia* where approximately 1% of users are contributing the majority of the content while 9% of users are usually editing, proofreading and correcting grammatical errors [27]. Finally, the majority of individuals, the remaining 90%, are only interested in using the available material without providing data or contributing to the overall efforts. In one of the empirical studies of this phenomenon, this observation was supported within the digital health social network context [28].

An approach to boost volunteer participation and user engagement comes through the gamification of the process [29–31]. The two mobile applications that were developed to utilize elements of gamification [32] but also support the data collection for the monitoring of the LC/LU and the collection of river parameters, are presented here.

2.3.1. Scent Explore

Scent Explore, Figure 3, is a game designed for the primary purpose of collecting scientific data and not for pure entertainment that the volunteers use to locate the PoIs, as defined by scientists through the *Campaign Manager*; carry out the tasks specified and gain rewards when successfully concluding them.

First, volunteers select a campaign and see the associated PoIs and the actions to be performed. As the volunteer reaches a PoI where an image of LC/LU is needed, the phone camera is activated and features of augmented reality, a *little animal*, are shown; see Figure 3. The volunteer is encouraged to look around and locate the *little animal*. While the volunteer is looking for the AR feature, the application is integrating the information received from the GPS with the gyroscope and the accelerometer so as to guide the user to the correct position, the one provided by the local authorities at the *Campaign Manager*. The little animal is carefully integrated to the screen so as to appear further away or close depending on the distance of the volunteer from the PoI. For older devices that lack some or all of the needed sensors for the navigation to the PoI, there are additional features that ensure an unobstructed user experience and valid collection of data. After locating the augmented reality feature, the volunteer can simply tap on the screen to capture it and as a result take an image with specific orientation that includes the information of interest [33,34]. The volunteer is then asked to choose elements of the taxonomy that best describe the content of the image. Upon the conclusion of this task, the user is rewarded by adding the captured *little animal* to their collection and with points depending on the distance from the PoI. Collecting points and badges, based on the level of expertise, is a common way of user engagement for gaming applications that has been proven to have positive results [35,36].

When a volunteer reaches a PoI where the water level measurement is required, the process described above is repeated, only this time the augmented reality elements are strategically located towards the water level indicator. Additional information about the image processing, its requirements and the measurement extraction are given in Section 2.4.1. For a PoI where the water velocity is required, the camera of the device is activated to video mode. The volunteer is then asked to capture a

video. Additional information about the requirements for the video and the extraction of the water surface velocity are provided in Section 2.4.2.

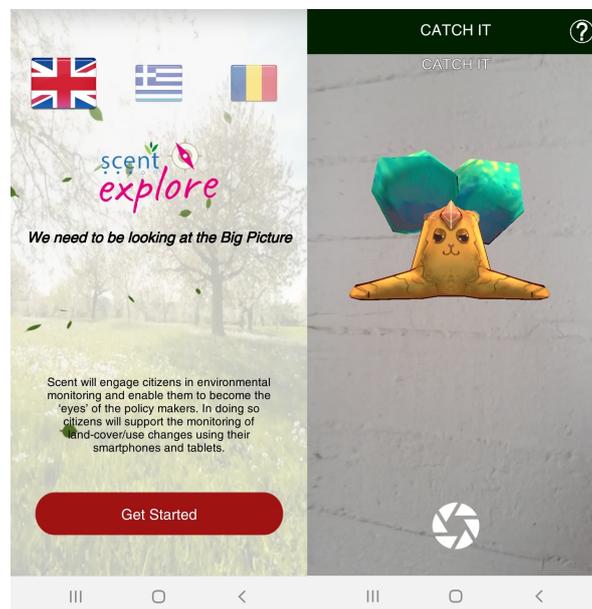


Figure 3. Scent Explore.

An additional feature of *Scent Explore*, aiming to improve the user experience and support user engagement, is the random PoI generator. In case that two PoIs, as defined in the *Campaign Manager*, have a significant distance or the user is delayed in any other way, e.g., people in the team taking longer to complete the PoI resulting in the user being inactive for more than three minutes, then the application creates a intermediate PoI requesting LC/LU information. This feature not only succeeds in keeping the volunteer engaged in case of delays during the campaign but also results in the collection of additional data.

Given that the collected video and images are further processed to extract river measurements, as discussed in Section 2.4, it is important they are of high quality. To this end, further actions are taken to ensure the quality of the collected multimedia. To begin with, images are stabilized by directly accessing the charge-coupled device camera sensor, improving the color schema and eliminating noise from movement, instead of using pictures interpolation. Image stabilization is additionally improved by taking three photos every time that a volunteer is capturing an augmented reality feature, rather than the expected one, and choosing the photo with less variation of the mobile accelerometer, to ensure better definition of the captured elements. Furthermore, the video stabilization is achieved by utilizing the accelerometer to detect instability and up-down movements at each frame. Additionally, extra thought is given to the positional data quality. The application gets GPS data with an implementation of Position Dilution of Precision (PDOP). This is very important given that with the GPS, optimal accuracy is in the order of 3–5 m or 10–20 m in isolated environments; with PDOP, the position accuracy can be less than 2 m or 3–5 m, respectively.

2.3.2. Scent Measure

The first step for the *Scent Measure* was to identify a set of requirements to be used in order to identify the proper sensor for the volunteers to use in the field. The requirements take into consideration the needed volume, the inexperience of the volunteers as well as the need for an interoperable and reproducible solution for different pilot areas. The identified requirements are presented in Table 1.

After an exhaustive search for the available sensors that can measure air temperature and soil moisture, the *Xiaomi Flower Care Smart Monitor* [37], as presented in Figure 4, was selected as it complies

with all the above mentioned requirements. It records measurements quickly and is simple to use as it does not require any installation. It provides measurements for air temperature and soil moisture every 15 s requiring only to insert the metal probes into the soil. It is a reliable sensor, with a battery life of a year and resistant to corrosion with a waterproof battery compartment. It is lightweight, only 130 g, and compact, only $12.00 \times 2.45 \times 1.25$ cm in dimensions, making it portable. The detectable temperature range covers from -20 °C to 50 °C with ± 0.5 °C temperature error control. The moisture is measured in scale 0%–100% with 1% accuracy.

Scent Measure, shown in Figure 5, is the application dedicated to the collection of sensor measurements by the volunteers. The application can be used independently, in areas where only sensor measurements are needed or complimentary to *Scent Explore*. The users can use their Scent account with this application too, in order to collect points and rewards for their contributions. In order to maintain the same look and feel among the applications, the user is rewarded with points for each collected measurement. In addition, badges dedicated to the collection of sensor measurement are available and unlocked based on the performance of the user.

The main data flow of the application is to connect to the portable sensor of the volunteer, collect measurements, add the needed metadata and forward them to the dedicated backend server. In detail, first the user can choose to login or play as a guest. Next, the application scans for nearby sensors. For communication with the portable sensor, the application uses Bluetooth Low Energy as it is designed to connect devices with low power consumption. All the discovered sensors are shown to the user so as to select the one that is assigned and proceed with the measurement. The application then collects air temperature and soil moisture values at 15secs time intervals from the selected sensor; the GPS coordinates and the current time as it is available at the mobile device are added as metadata to the measurement which is then forwarded to the backend server. To further support user engagement, the application notifies the user every time a new measurement is recorded and displays the recently collected measurements.



Figure 4. Xiaomi Flower Care Smart Monitor.

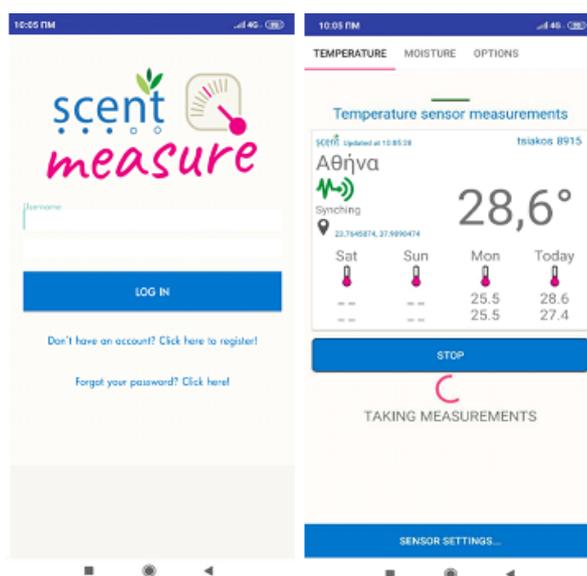


Figure 5. Scent Measure.

Table 1. Requirements for portable sensor.

Requirement	Justification
Measurements to be recorded.	According to the requirements of the environmental scientists, the key parameters that the portable sensors should be able to record are air temperature and soil moisture.
Low-cost.	As many sensors as possible are needed in the field simultaneously; where it is probable they might get destroyed or lost, the price per unit for the sensor should be low. This, however, should not compromise the quality and accuracy of the collected measurements.
Portability.	Keeping in mind that the volunteers will have to carry the sensors for the duration of the Campaign, there is a need for the sensor to be lightweight. In addition, it is crucial to ensure that the sensor will have enough power autonomy for at least one fully operational day, to avoid the need for power banks or charging stations during the Campaign.
Ease-of-use.	Taking into consideration the inexperience of the volunteers, their lack of training regarding the scientific way for the collection of environmental measurements, the possible challenges on the field and the need for multiple measurements in a short time span, the portable sensor should have a very intuitive way of use and require minimum effort from the user.
Connectivity.	Aiming to have a high degree of accuracy and ensure uniformity among the measurements collected, many of the needed values such as the GPS coordinates and the timestamp are collected from the mobile device of the volunteer. This means that the sensor should be able to connect, preferably over Bluetooth, with a mobile device and forward the information there.
Open API.	Last but not least, the collected information is to be forwarded to the system backend. In order for that to be feasible the sensor should provide an open API.

2.4. River Measurement Extraction from Multimedia

The collection of measurements from water bodies, such as rivers and canals, are usually carried out by groups of scientists trained to accurately collect such measurements without bias, using a variety of methods to ensure quality control [38–40]. Such data collections require expensive equipment and are time consuming. The utilization of such techniques in the context of citizen science is not feasible as they are often complex, require training and resources and lead to variable degrees of data quality based on the expertise of the volunteer contributing.

In order to provide measurements for the proper monitoring of ecosystems, the data should be consistent, accurate and independent of the expertise of the volunteers collecting it. To achieve this, volunteers are simply asked to take a picture of a water level indicator or a video of a predefined floating object and upload it through the *Scent Explore*. The crowd-sourced data are then processed by two tools, that have been specifically designed to automatically extract the water level and the water velocity in a consistent and accurate way.

2.4.1. Water Level Measurement Tool

The WLMT receives an image that contains a water level indicator and the waterline, the line where the water level indicator meets the surface of the water. Initially, the tool uses the color difference in the image to identify the waterline. This step is very important for the detection of the correct measurement, the numbers closer to the waterline, and eliminates any noise from numbers of the water level indicator higher up or random signage included in the image.

Next, the tool “reads” the indicator and extracts the number that is closest to the waterline. The conversion of images containing handwritten or printed text into machine encoded text is called Optical Character Recognition or Optical Character Reader (OCR). Tesseract [41] is an optical recognition engine, released under the Apache License, Version 2.0, and sponsored by Google since

2006 [42]. Tesseract is considered one of the most accurate open-source OCR engines available [43,44]. In our case, a Python script utilizes this library to extract the water level.

While the digit extraction is very accurate and robust, the main issue is the differentiation between water level indications, specifically if a topological or a hydrological indicator is captured in the image. In the majority of the campaigns, hydrological indicators were used, but in some cases during the first campaigns where the issue was not known, some topological indicators were used as well. This is an important problem as the two indicators display the numbers very differently. The hydrological indicators use centimeters while the topological indicators use decimeters, so an identified value of 10 corresponds to 10 cm for a hydrological indicator but to 10 dm for a topological one.

Aiming to ensure measurements with high degree of confidence, detailed instructions are given to the volunteers, as presented in Table 2, regarding the optimal way that they are to capture images including a water level indicator.

Table 2. Instructions for the water level indicator capture.

Step Number	Description
#1	Locate the water level indicator at the opposite river bank.
#2	Wait for the camera to be activated.
#3	Stand straight opposite to the water level indicator.
#4	Chose the camera frame to include the water level indicator and the waterline.
#5	Make sure that you keep the camera stable and steady.
#6	Capture the image.

2.4.2. Water Velocity Calculation Tool

The WVCT receives a video with a pre-defined object, which is floating downstream following the river course, and extracts the water surface velocity. As a floating object, a tennis ball was selected. Initially, efforts were dedicated to more eco-friendly solutions such as oranges. Such solutions however pose significant technical challenges, given that each object is expected to have different color and size and differ in density. Upon testing, it was observed that occasionally oranges were submerged into the water rather than floating on the water surface. Tennis balls have strict specifications, they are always the same bright yellow color, have a 6cm diameter and float on the water surface. To ensure that no tennis balls will pollute the ecosystem, a very thin fishing string was carefully attached to the tennis balls used during the campaigns to ensure that the ball would be always retrieved.

The main challenge regarding the processing of the video is the noise that has been introduced to the recording due to intentional or unintentional movement of the mobile device. In order to eliminate this noise the video is first stabilized. The stabilized video is then processed frame by frame. The first step is to locate the first frame of the video that contains the tennis ball. This is achieved by looking for the very distinct bright yellow color of the tennis ball, which is not expected to occur naturally in the environment. Each pair of consecutive frames are examined, and the optical flow between them is calculated. First, the average optical flow of the area of the frame that includes the tennis ball is calculated. Next, the average optical flow of the rest of frame is calculated to identify if the camera was moving during the video capture and eliminate this movement from the calculation. The displacement of the ball, between these two frames, is calculated by subtracting from the average optical flow of the ball the average optical flow of the rest of frame.

The calculation is repeated as long as the next video frame contains the tennis ball. A dedicated counter is recording the number of frames that were valid and examined while the total displacement is calculated as the sum of the displacements between all pairs. The end of the video is defined as the first frame where a tennis ball cannot be located. This is an additional safety measure in the case where excessive movement of the camera resulted in the tennis ball being completely removed from

the frame. The total displacement, which has been calculated in pixels, is then calibrated using the dimension of the tennis ball given, which we know to be 6 cm in diameter. Finally, the number of frames where the tennis ball is found is calibrated by the frames per second of the video, to give the time in seconds calculated for the displacement. We calculate the water velocity by dividing the total displacement by this time resulting in the measurement of the water surface velocity in cm/s.

The tool has some baseline validation mechanisms that are used to discharge video that is of low quality. The first validation test is with regard to the duration of the presence of the tennis ball. If the tennis ball is tracked in fewer frames than the frames per second of the video, giving less than a second of useful material, then the video is considered invalid. The second validation test examines the calculated displacement of the tennis ball. If the displacement is less than the identified size of the tennis ball, then the video is again considered invalid. In any other case, the water velocity is extracted with a confidence score that it calculates as the average of the percentage of the video frames that were used, continuous identification of the tennis ball, to the total duration of the video and the average number of pixels that the tennis ball covers in the frames to the overall displacement as calculated in pixels. The function giving the confidence level is:

$$ConfidenceLevel = \frac{1}{2} \left(\frac{ValidVideoFrames}{TotalVideoFrames} + \frac{TennisBallSize}{Displacement} \right)$$

The tool was tested with some videos captured for this purpose, containing frames of the tennis ball. The tracking of the object was very reliable and robust to the video quality and duration; the displacement calculation however was up to some degree dependent on the stability of the video. Aiming to ensure measurements with high degree of confidence, detailed instructions are given to the volunteers, as presented in Table 3, regarding the optimal way that they are to capture the video.

Table 3. Instructions for video recording.

Step Number	Description
#1	Select an area of the river where there are no visible obstacles and/or river bends.
#2	Activate the camera and chose the appropriate frame.
#3	Throw the securely attached tennis ball upstream.
#4	Start the video recording while waiting for the object to float by.
#5	Record the object floating by while keeping the camera stable and steady.
#6	When the object is no longer visible within the camera frame stop the video recording.
#7	Retrieve the tennis ball using the fishing line.

3. Demonstration Cities and Data Collection

The Scent Toolbox was tested in two carefully selected areas, the Danube Delta in Romania and the Kifisos River basin in Greece. The areas offer different environments and topologies and have different needs and challenges allowing the testing of the system under different conditions. Starting from August 2018, several specific campaigns were organized for each study area. Each campaign was focused on one of the needed measurements in order to facilitate a training workshop before each one. During the workshop, the volunteers were informed about the project and trained in the use of the Scent Toolbox component that was used during the campaign. The campaigns ran for a period of 10 months, with 6 different campaigns taking place in each pilot area, two for each thematic topic, LC/LU images, sensor measurements and river data.

3.1. Danube Delta, Tulcea, Romania

The Danube Delta in Tulcea, Romania [45] is the largest natural wetland in Europe and is protected under UNESCO as a unique biosphere reserve. The Danube Delta Biosphere Reserve has the third largest biodiversity in the world with more than 5500 species of flora and fauna and the highest concentration of bird colonies in Europe. Many of the species of plants and animals living in the Danube Delta are unique to it. The Biosphere Reserve is also home to a population of more than 10,000 people, who rely on the natural resources and the environment of the Danube Delta for their livelihoods.

However, the Delta has suffered from human interventions, which led to dramatic changes in the area, including damming large areas for agricultural use, fishing and forestry. This has resulted in the disturbance of the ecological balance of the wetlands and in some cases the deterioration and loss of biodiversity. Monitoring the changing landscape of the Delta is an important first step in maintaining its natural ecological balance, and protecting its communities. The Scent Toolbox aims to collect up-to-date data to dynamically and accurately monitor changes in land-cover and land-use and thereby help to better protect both the Delta's environment and the population who live there.

Campaigns in Numbers

During the campaigns 200 participants collected 1500 videos, 1500 sensor measurements, 1800 images containing water level indicators and 16,000 images with LC/LU elements.

3.2. Kifisos River Basin, Attica, Greece

The Kifisos river basin [46] is roughly 380 km², and almost 60% of its watershed is urbanized as the metropolitan area of Athens. As the city has expanded, the land-cover of the area has transitioned from rural to urban, and industrial in some areas. The hydrological network of the basin has been heavily engineered to support expanding constructions. However, in many cases, the hydraulic works were poorly designed. There are many areas where there are illegal constructions, even within the main river course. As a result, during periods of heavy and rapid rain events, the river floods due to the insufficiency of drainage networks, causing severe damage to infrastructure around the river.

Scent has organized campaigns in areas of the river basin characterized by rapid changes in the river water level caused by natural and anthropogenic factors. With the Scent Toolbox, volunteers collected important information that allows for increased awareness of the status of the river environment. The volunteers have also collected soil moisture and temperature measurements, both before and after heavy rain events to help hydrologists better understand the river dynamics.

Campaigns in Numbers

During the campaigns 460 participants collected 400 videos, 1000 sensor measurements, 700 images containing water level indicators and 4000 images with LC/LU elements.

3.3. Safety Considerations during Campaigns

For the campaigns organized at the Danube Delta, personnel from the DDNI [47] and SOR [48] were responsible for ensuring the safety of the volunteers. Both organizations are partners of the Scent project and experienced in organizing campaigns for educational or bird watching purposes in the area that was visited. In all cases, safety precautions were taken, including life vests for boat trips, protection from the weather elements such as hats and sunscreen in the summer and raincoats in the winter as well as satellite and radio communications in case of an emergency while at remote areas without GSM (Global System for Mobile Communications) signal.

For the campaigns organized at the Kifisos river in Attica the Hellenic Rescue Team of Attica (HRTA) [49], a Scent project partner was responsible for the safety of the volunteers. The HRTA members have significant experience in providing first-response emergency medicine in events with many participants as needed. For most of the campaigns organized at the Kifisos river, pedestrian paths

in semi-rural environment were chosen, so the HRTA members were mostly tasked with coordinating the teams of volunteers and ensuring that all safety guidelines were respected. As an exception to this, due to the topological challenges of an area of interest at the northern part of the Kifisos river, only trained HRTA members were allowed to participate.

4. Data Quality Control

4.1. Lc/Lu Annotations

Volunteers are given training over the available taxonomy elements as well as examples of images to which these correspond. This training process provides the relevant knowledge to the volunteers in order to provide annotations of high accuracy. The environment of the campaign, the high inter class variation of the taxonomy and the unpredictability of the environment may lead the volunteer to provide incorrect or misleading annotations. In order to ensure that such annotations will be removed from the dataset, a validation mechanism has been implemented.

Scent Collaborate is a browser-based crowd-sourcing platform that allows users to validate images collected and annotated during the campaigns. The users can also provide further annotations for elements that are in the images but have not been included in the annotations. Each image is annotated by one or multiple users, depending on whether there is an agreement or not upon the validity of the annotation. The validity of the annotation is established based on the majority vote.

Aiming to boost user engagement and the user-friendliness of the toolbox, the user can login to the Scent Collaborate using the Scent account. Then the user is rewarded with points based on the number of annotated pictures. Once more, the points are connected to badges that acknowledge the achievements of the users. Scent Collaborate is designed to engage users that do not live close to any of the study areas but are still interested in contributing to the Scent movement.

4.2. River Data Measurements

Understanding that due to low quality contributions and calculation inaccuracies, the measurements extracted by the WLMT and the WVCT tools can contain invalid data, a methodology has been developed that identifies such values and removes them from the datasets. We present here the steps of the methodology as well as the results of the data quality analysis for the data collected during the first Kifisos campaign, held from the 15th to the 17th of November, 2018.

The first step for the evaluation of the data extracted by the WLMT is the spatial and temporal clustering. To this end, the coordinates and timestamps available from the metadata were used to divide the measurements into K clusters using the K-means algorithm [50,51]. The main challenge of the k-clustering process is to correctly identify the number K of the clusters into which the data should be divided. While knowing the nature of the data and the collection process can provide an intuitive way of specifying the number of clusters, we chose to use a more robust and agnostic approach, the average silhouette method [52]. Applying this step to the first Kifisos campaign, the method indicated that the optimal number of clusters is $K = 10$; Figure 6. Further examining the clustered data, an internal variance was observed along with data that can be considered as outliers. In Figure 7 at the top, the internal variance of the clusters is presented using the box-plot technique. The next step of our methodology was to eliminate the internal variance presented in the clusters. To achieve that, the “sigma test” was applied [53].

This test checks if data within the same cluster follow the normal distribution $N(\mu, \sigma^2)$, and the mean value and standard deviation can be approximated by the characteristics of the sample, $\mu \approx \bar{x}$ and $\sigma^2 \approx s^2$. The sigma test is used to remove possible invalid measurements by keeping only the data in the interval $(\mu - c \cdot \sigma, \mu + c \cdot \sigma)$, with $c \in \mathbb{R}^+$.

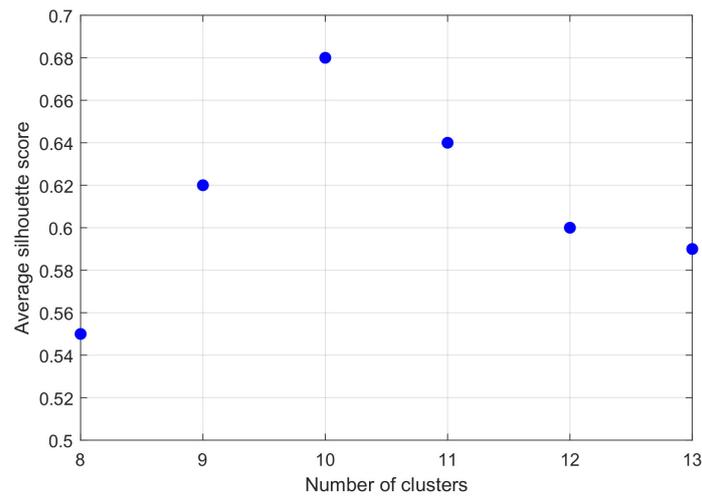


Figure 6. Using the silhouette method to define number of clusters.

For the data extracted during the Kifisos campaign, and for a value $c = 1$, about 32% of data of each cluster were invalidated. The box-plots of the clusters after the sigma test are presented in Figure 7 at the bottom. Carefully examining the results, it is noticeable that some variance is still within the results. In order to identify the source of this variance, the outliers were visually inspected. This showed that occasionally the WLMT gave a wrong measurement. In most cases, the WLMT gave the proper reading based on the content of the image but that was significantly higher than expected as the bottom of the image did not contain the water level, or the water level indicator was obstructed by brought material. In a few limited cases, deterioration of the letters printed on the indicators made some digits resemble others such as a “7” resembling a “1” or an “8” resembling a “0”.

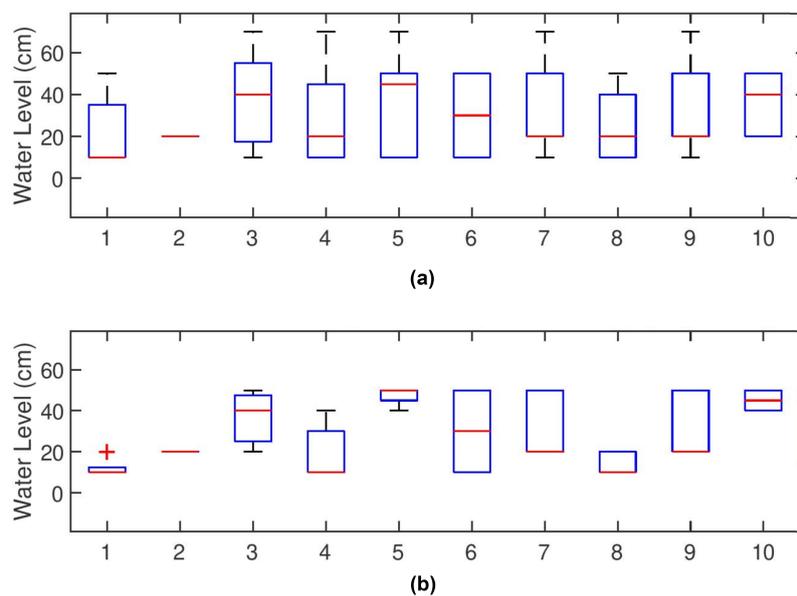


Figure 7. The variance of water level measurements (a) before and (b) after the sigma test.

For the evaluation of measurements extracted by the WVCT, the same methodology was followed. In Figure 8, the box-plots that show the range of the collected data of the $K = 12$ clusters for the Kifisos campaign are present. The number of clusters was once more decided using the average silhouette method. In some cases, there are clear outliers; these were examined further to identify the causes. The main cause was that often part of the video contained the retrieval of the tennis ball, a process that was increasing the average velocity calculated. In order to eliminate this problem, instructions

were given to the volunteers and an option was added to the interface of the Scent Explore application allowing users to discard video that was not properly captured.

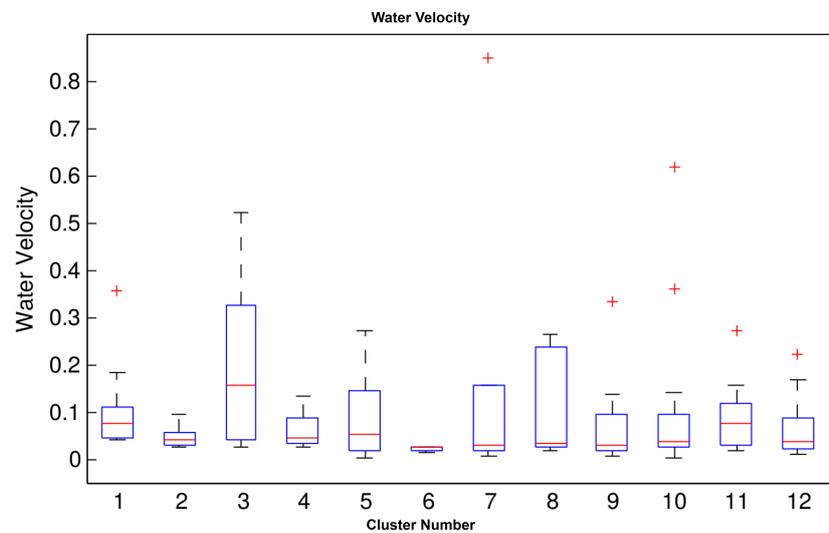


Figure 8. The water velocity data divided in $K = 12$ clusters.

4.3. Sensor Measurements

Sensor measurements are accompanied by rich metadata that include coordinates, timestamp, a unique measurement ID, the unique identifier of the volunteer that collected the measurement as well as the GPS accuracy. The measurements contain two values, one for air-temperature measured in Celsius degrees and one for soil moisture as a percentage. The collected data of each campaign are examined to identify potential invalid measurements and remove outliers. For both of the collected measurements, the criteria used to identify invalid measurement are

- Spatially isolated measurement. Knowing that volunteers collect measurements in groups of at least three persons, having a measurement that is more than 10 m away from the rest is an indication of either invalid coordinates or a contribution far from the PoI.
- Volunteer exclusive measurements. This test is also based in the knowledge of how campaigns are organized, in groups of people. If there is an area where only one volunteer is providing measurements, most probably there was an issue with the provided coordinates or the execution of the measurement collection process.
- GPS accuracy being greater than 20 m. This test is necessary, as the areas the volunteers visited were remote, and in many cases the GPS coordinates were not accurate. Given the importance of the spatial distribution for the usage of the collected measurements, such values are not useful.

For the air temperature measurements, an additional validation test, called the range test, is carried out. The validate range is calculated as $\bar{T} \pm 10$ °C, where \bar{T} is the daily mean temperature the volunteers provided. For the soil moisture measurements, the delta test [53] is applied. The analysis of the soil moisture measurements, revealed many entries with zero as value. Extensive laboratory testing proved that the sensor can provide a valid measurement of 0% moisture level when measuring dry soil without any roots of living plants inside, which was also expected by its specifications. This created the need for differentiating between measurements taken with the sensor not inserted properly to the ground thus being invalid and measurements taken where the soil was dry. In order to identify valid zero measurements, the time series of measurements provided by a user are examined. If a 0% moisture level measurement is followed within the next 15 s, which is the update rate of the sensor, with a measurement greater than 0%, then the first measurement is invalidated.

For the Kifisos campaign, there were $N = 215$ measurements of air temperature and soil moisture collected. Following the methodology described before, they were clustered into $K = 4$ clusters.

The spatial distribution of the measurements collected is presented in Figure 9, showing that the choice of creating four clusters was accurate. The internal cluster distribution of the air temperature is uniform, distributed inside its range in most cases. However, there are cases where the distribution of the values within a cluster is irregular; such an example is presented in Figure 10 where the data distributions within two clusters are compared. This raises the question of a user providing systematically invalid measurements.

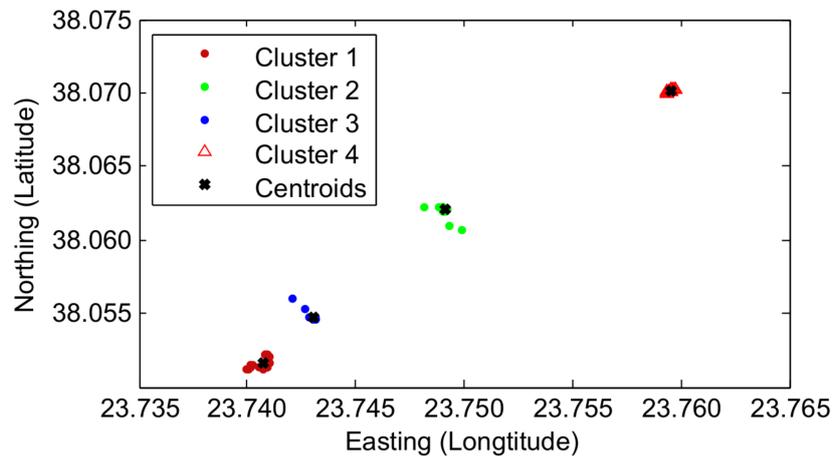


Figure 9. Spatial distribution of the sensor data clusters.

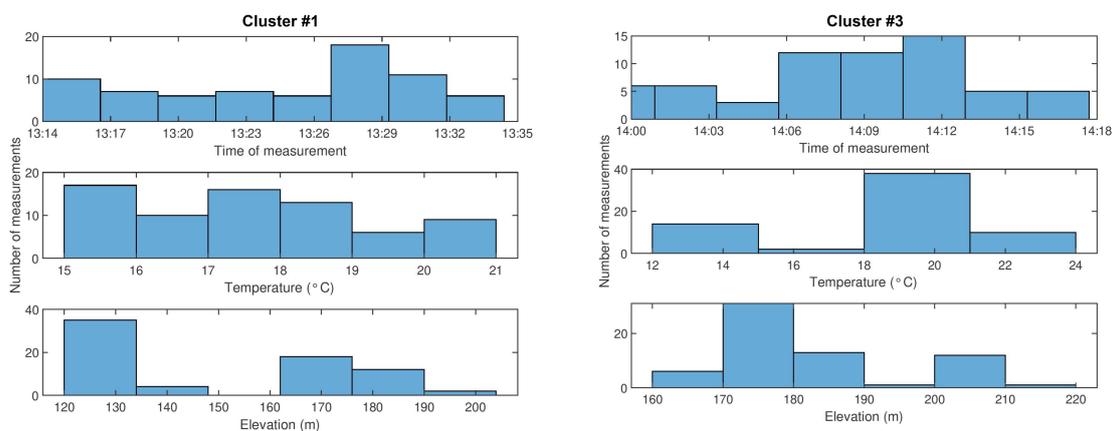


Figure 10. Normal vs irregular intra-cluster data distribution.

In order to determine that, the data of the cluster are further divided using the unique user identifier, as shown in Figure 11, and the one-way ANOVA statistical test [54,55] is applied as shown in Table 4. This test proves that there is significant difference between measurements collected by different users within the cluster, as $p = 6.47 \times 10^{-34} < 0.01$. The mean air temperature values showed that the user with unique identifier 368 consistently recorded higher temperatures.

Table 4. Data evaluation in cluster using the ANOVA table.

Source	SS	df	MS	F	Prob > F
Group	499	3	166	251	6.47×10^{-34}
Error	39	60	0.66		
Total	539	63			

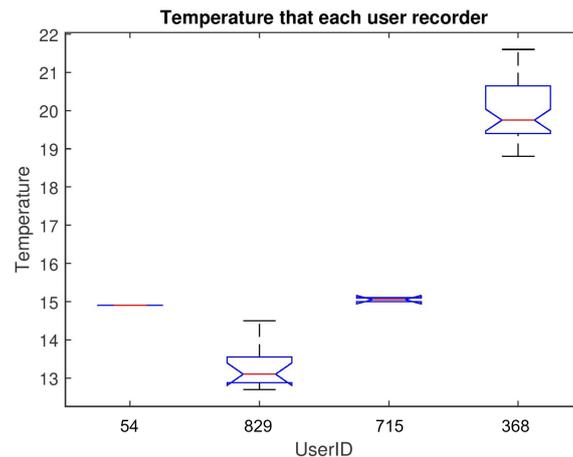


Figure 11. Intra-cluster data grouped by user ID.

5. Data Provision

Taking into consideration the large volume of collected data, the need of the environmental scientists and local authorities for information based on spatial and temporal criteria as well as the importance of interoperability and data discovery, the OGC SensorThings API [22] is implemented as the storage and retrieval platform for all measurements collected [56]. The SensorThings API is an OGC standard providing an open and unified framework to interconnect sensing devices with data and applications. It follows REST principles, the JSON encoding, and the OASIS OData protocol and URL conventions.

The foundation of the SensorThings API is its data model that is based on the ISO 19156, ISO/OGC Observations and Measurements, which defines a conceptual model for observations and for features involved in sampling when making observations. In the context of the SensorThings API, the features are modeled as Things, Sensors and Features of Interest.

5.1. Data Harmonization

Each data source provides the data in custom JSON formats, and these are received by the Data Translators, the entry points of information for the back-end. Each translator receives data from a specific source and pushes the data to the Data Quality Control modules as needed. The Data Translator also ensure that the received data respect the format and types of the data models. They use the available information in the received data object, trying to transform it based on the requirements of the SensorThings API standard. In case the data received are not complete or the JSON values are not the proper data types, they discard the data object, properly logging this information. In any other case, they forward the transformed data objects to the Data Quality Control module.

For the SensorThings API, the correlation between the received data and the data model involves the creation of two or more data objects for each received measurement. For each data object received, an Observation and a FeatureOfInterest are created containing the measurement collected and the location that is associated with it. In order for these entities to be valid, in the context of the standard, an Observation should be associated with a Datastream. A Datastream is a unique combination of a Thing, a Sensor and an ObservedProperty. Here, the pre-defined the observed properties based on the expected data are presented:

- Water level;
- Water surface velocity;
- Air Temperature;
- Soil moisture;
- LC/LU.

Each volunteer is modeled as a Thing, based on the unique user identifier that is available in the received metadata. Each portable sensor is modeled as a Sensor, based on the sensor unique identifier included in the metadata. For the data that do not require a portable sensor, the mobile device of the volunteer, uniquely identified by the unique user identifier of the volunteer, is used as a Sensor.

5.2. Data Storage

Given the need for a scalable storage solution that can support the provision of spatially dense data with variant time distributions, without a known volume and the need to efficiently respond to spatial and temporal queries from multiple users simultaneously, a distributed storage schema is used for SensorThings API data schema.

The majority of the databases nowadays offer some support for spatial queries. Most of the traditional approaches such as the ones in MySQL [57] and PostGIS [58] are based on R-tree indexes to respond to spatial queries. Such approach has a key challenge, closely related to the spatial distribution of the stored information. R-trees are only effective if they are balanced, meaning that the stored rectangles do not overlap or have empty spaces, as they do not guarantee good worst-case performance. The quality of the R-tree will determine the number of sub-trees that will be examined each time to answer a query. In our case, given that all the measurements are collected at or near PoIs, it is expected that there will be multiple measurements with the exact same location. The lack of spatial deviation means that it is close to impossible to construct a balanced R-tree.

Another technique for spatial indexing is the Z-ordering, which aims to alleviate the problems of the R-tree index. The Z-ordering is based on a transformation of the input spatial information to a one-dimensional data structure without any loss of information. While, this approach can handle efficiently spatial points, it lacks the flexibility to efficiently store multiple overlapping objects that lack spatial distribution.

The XZ-ordering [59] is an index that can efficiently handle the lack of spatial distribution. This is achieved by avoiding object duplication and allowing overlapping Z-elements. To this end, in order to ensure the optimal storage and indexing of the data, the XZ-ordering index for the spatial information was chosen for the SensorThings API. This is achieved by using the GeoMesa [60] indexing tools over an Accumulo [61] datastore.

5.3. Ogc Sensorthings Api

We have implemented the SensorThings API as a standalone Tomcat application using the Jersey RESTful Web Services framework. The implementation supports all the requests described in the standard as well as all the filtering capabilities. The implementation has received the compliance badge recognizing that it is an OGC certified product [62]. The endpoints provided are presented below.

5.3.1. GET Requests

Entities may be accessed in many different ways:

- As a collection of entities: A list of all entities in the specified entity set when there is no service-driven pagination imposed. The response is represented as a JSON object containing a name/value pair named value, a JSON array where each element is a representation of an entity or of an entity reference.
- By identifying one entity in a collection: A JSON object of the entity that holds the specified ID in the entity set.
- By identifying a property of a single entity: The specified property of an entity that holds the ID in the entity set.
- By accessing the value of an entity's property: The raw value of the specified property of an entity that holds the ID in the entity set.

- By following a navigation property: A JSON object of one entity or a JSON array of many entities that holds a certain relationship with the specified entity.
- By following an association link: A JSON object with a value property, a JSON array containing one element for each associationLink. Each element is a JSON object with a pair, with name URL and the value of selfLinks of the related entities.

5.3.2. POST Requests

To create an entity in a collection, the client sends a HTTP POST request to that collection's URL. The POST body contains a single valid entity representation. If the target URL for the collection is a navigationLink, the new entity is automatically linked to the entity containing the navigationLink.

Newly created entities may be linked to existing entities, by using the entity ID, or contain the information needed to create the related entities. A request to create an entity that includes related entities, represented using the appropriate inline representation, is referred to as a "deep insert".

5.3.3. PATCH Requests

HTTP PATCH requests are responsible to merge the content in the request payload with the entity's current state, applying the update only to those components specified in the request body. The properties provided in the payload corresponding to properties that can be updated replace the values in the entity. Missing properties of the containing entity or complex properties are not directly altered. These requests may contain binding information for navigation properties. For single-valued navigation properties, this replaces the relationship. For collection-valued navigation properties, this adds to the relationship.

5.3.4. DELETE Requests

A successful DELETE request to an entity's edit URL deletes the entity. The request body must be empty. The implementation implicitly removes relations to and from an entity when deleting it; clients need not delete the relations explicitly. Related entities may be deleted or modified if required by integrity constraints.

6. Discussion of Similar Approaches

According to relevant studies [63], the first recorded effort of establishing a Citizen Science project is from 1989, when 225 volunteers across the US were asked to collect rain samples to assist the Audubon Society in an acid-rain awareness campaign. The volunteers collected samples, recorded pH, and reported back to the organization. The information was then used to demonstrate the full extent of the phenomenon [64]. While this project is very important as it established the advantages of engaging volunteers in scientific projects, the data collection process was far from simple; the sample pool was very limited, and the data collected were not validated in any way, making this a poorly designed Citizen Science project by today's criteria.

Since then, many Citizen Science projects have been established, from different scientific areas, aiming to collect contributions from volunteers. To the best of our knowledge, SCENT is the only Citizen Science project tackling the issue of monitoring water parameters such as water level and velocity in the context of aquatic ecosystems, so a direct comparison of the results is not feasible. The design decisions, the system architecture, the validation process as well as the campaign organization will be compared here with other Citizen Science projects that are focusing on monitoring other elements of the environment.

A very interesting approach was the bumblebee [65] monitoring in the UK that was supported by citizen science. Data were collected on 1022 bumblebee nests when participants were asked to record attributes of bumblebee nests discovered in their gardens and either fill an online form or provide their data with post. The collected data were very valuable, as they made it feasible to study how

bumblebees select their habitat, and provided the first quantitative evidence of potential decline in one of the UK's 'big six' common bumblebee species. This was possible due to the wide geographic range in a short time period of the data collection. The main challenge this project faced was the identification of the species. This required photographs that were examined by experts. This is a major scalability issue as an increase in the number of collected data samples would make their validation impractical.

eBird [66] is a program launched by the Cornell Lab of Ornithology (CLO) and the National Audubon Society in 2002, which engages a vast network of citizen scientists in reporting bird observations using standardized protocols. The mission is to better understand bird distribution and abundance across large spatiotemporal scales and to identify the factors that influence bird distribution patterns. The collected information provides useful insights on species occurrence, migration timing, and relative abundance at a variety of spatial and temporal scales. To effectively engage birders, eBird provides a permanent repository for their observations and a method for keeping track of each user's personal observations, birding effort and various bird lists. The main challenge of this project is that data are collected by individuals and not by organized campaigns. Inexperienced users might use bird counting techniques wrongly or confuse similar looking bird species. If these users are in locations where not enough information is available, bias may be introduced to the system.

CoralWatch [67], launched in 2002, is a citizen-science program that seeks to integrate education and global reef monitoring by examining coral bleaching and uses a monitoring network to educate the public about reef biology, climate change and environmental stewardship. The main tool for CoralWatch participants is a square of plastic that can be used like a color swatch card, to monitor coral bleaching and therefore coral health. CoralWatch participants match the chart colors to the coral color, record the codes and enter the data on a chart via a website. Scientists developed the colors on the chart using intentionally bleached coral in temperature-controlled aquaria. In this case, the data collection protocol is the main challenge. In order for the information collected to be of use to the scientists, additional data to the color information is needed, putting the user engagement at risk. Special effort was made by the project organizers to communicate with participants through various media.

The plastic pollution is a fast escalating problem for all the oceans around the world. Careful monitoring and recording of their spatial and temporal distribution, a key requirement to identify the origin and measures to contain the pollution, is a major and expensive task that requires a vast amount of properly distributed data. The citizen science project "National Sampling of Small Plastic Debris" [68] was supported by schoolchildren from all over Chile who documented the distribution and abundance of small plastic debris on Chilean beaches. Thirty-nine schools and nearly 1000 students from continental Chile and Easter Island participated in the activity. To validate the data obtained by the students, all samples were recounted in the laboratory, again raising the issue of scalability with regard to the validity of the data.

In all the Citizen Science projects presented here, the main issue was the data validity. In most cases, experts were tasked with the manual inspection of the collected information or even the duplication of the measurement in order to ensure a high quality of data. Such approaches, however, create the question of the scalability and the reproduction of the experiment in different areas or countries interested in participating. The Scent toolbox, through dedicated tools and applications, has automated the data validation process, creating a scalable solution that can be easily reproduced in other areas of interest.

7. Conclusions

We have presented here an innovative monitoring system and a constellation of technologies that can support the participation of volunteers in the collection of data for environmental ecosystems. Carefully designed tools, easy to use and reliable sensors along with a thorough data quality mechanism allow citizens to contribute proper, unbiased and high quality scientific data as needed for the researchers and local authorities. The collected data are harmonized and offered to environmental scientists and local authorities allowing a more targeted monitoring of areas of interest.

Author Contributions: Software, M.K., S.-N.B., T.T., S.T.; validation, A.R.; writing—original draft, M.K., L.S., A.R.; writing—review and editing, M.K.; supervision, E.S.; project administration, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement no 688930, project SCENT (Smart Toolbox for Engaging Citizens into a People-Centric Observation Web). Content reflects only the authors’ view and the European Commission is not responsible for any use that may be made of the information it contains. For more information about the SCENT project visit the website <https://scent-project.eu/>.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

API	Application programming interface
OGC	Open Geospatial Consortium
LC/LU	Land Cover/ Land Use
CORINE	Coordination of Information on the Environment
CLC	CORINE Land Cover
MMU	Minimum Mapping Unit
PoI	Point of Interest
WLMT	Water Level Measurement Tool
WVCT	Water Velocity Calculation Tool

Appendix A. Taxonomy Elements

The taxonomy elements used for the LC/LU monitoring are shown in Table A1.

Table A1. Taxonomy Elements.

Urban Case—LC/LU	Obstacles or Damages	Rural Case—LC/LU
Buildings	Trees	Furrows
Paved areas	Tree branches	Rice fields
River bank. Low grass	Cars/vehicles	Vineyards
River bank. Shrubs	Dustbins	Fruit trees
River bank. Concrete	Storm drains. Clean	Pastures/Grassland
River bank. Stone	Storm drains. Debris	Forest. Broad-leaved
River bank. Bare soil	Storm drains. Leaves	Forest. Coniferous
River. Dry cross-section	Storm drain discharge	Forest. Mixed
Roads	Buildings. Damaged	Heathland
Parking lots	Roads. Damaged	Sand
Railway lines		Bare rock
Excavation or construction sites		Sparsely vegetated area
Dump sites		Burnt area
Parks. Tall vegetation		Inland marshes

References

1. Silvertown, J. A new dawn for citizen science. *Trends Ecol. Evol.* **2009**, *24*, 467–471. [[CrossRef](#)] [[PubMed](#)]
2. Dickinson, J.L.; Shirk, J.; Bonter, D.; Bonney, R.; Crain, R.L.; Martin, J.; Phillips, T.; Purcell, K. The current state of citizen science as a tool for ecological research and public engagement. *Front. Ecol. Environ.* **2012**, *10*, 291–297. [[CrossRef](#)]
3. RCR. Data Collection Available online: https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dctopic.html (accessed on 19 April 2020).

4. Geoghegan, H.; Dyke, A.; Pateman, R.; West, S.; Everett, G. *Understanding Motivations for Citizen Science*; UK Environ. Obs. Framew: Swindon, UK, 2016.
5. Dickinson, J.L.; Zuckerberg, B.; Bonter, D.N. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annu. Rev. Ecol. Evol. Syst.* **2010**, *41*, 149–172, doi:10.1146/annurev-ecolsys-102209-144636. [[CrossRef](#)]
6. Conrad, C.C.; Hilchey, K.G. A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environ. Monit. Assess.* **2011**, *176*, 273–291, doi:10.1007/s10661-010-1582-5. [[CrossRef](#)] [[PubMed](#)]
7. Preece, J. Citizen Science: New Research Challenges for Human—Computer Interaction. *Int. J. Hum. Comput. Interact.* **2016**, *32*, 585–612, doi:10.1080/10447318.2016.1194153. [[CrossRef](#)]
8. Newman, G.; Crall, A.; Laituri, M.; Graham, J.; Stohlgren, T.; Moore, J.C.; Kodrich, K.; Holfelder, K.A. Teaching citizen science skills misc: Implications for invasive species training programs. *Appl. Environ. Educ. Commun.* **2010**, *9*, 276–286. [[CrossRef](#)]
9. See, L.; Comber, A.; Salk, C.; Fritz, S.; van der Velde, M.; Perger, C.; Schill, C.; McCallum, I.; Kraxner, F.; Obersteiner, M. Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. *PLoS ONE* **2013**, *8*, 1–11, doi:10.1371/journal.pone.0069958. [[CrossRef](#)]
10. Kosmala, M.; Wiggins, A.; Swanson, A.; Simmons, B. Assessing data quality in citizen science. *Front. Ecol. Environ.* **2016**, *14*, 551–560, doi:10.1002/fee.1436. [[CrossRef](#)]
11. Trench, B. Science communication and citizen science: How dead is the deficit model. In Proceedings of the IX International Conference on Public Communication of Science and Technology (PCST), Seoul, Korea, 17–20 May 2006.
12. Rotman, D.; Preece, J.; Hammock, J.; Procita, K.; Hansen, D.; Parr, C.; Lewis, D.; Jacobs, D. Dynamic Changes in Motivation in Collaborative Citizen-science Projects. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*; ACM: New York, NY, USA, 2012; pp. 217–226, doi:10.1145/2145204.2145238. [[CrossRef](#)]
13. Land-Zandstra, A.M.; Devilee, J.L.A.; Snik, F.; Buurmeijer, F.; van den Broek, J.M. Citizen science on a smartphone: Participants’ motivations and learning. *Public Underst. Sci.* **2016**, *25*, 45–60, doi:10.1177/0963662515602406. [[CrossRef](#)]
14. Murphy, C.; Wilby, R.L.; Tom, K.M.; Thorne, P.; Broderick, C.; Fealy, R.; Hall, J.; Harrigan, S.; Jones, P.; McCarthy, G.; et al. Multi-century trends to wetter winters and drier summers in the England and Wales precipitation series explained by observational and sampling bias in early records. *Int. J. Climatol.* **2019**, *40*, 610–619, doi:10.1002/joc.6208. [[CrossRef](#)]
15. Gupta, P.; Doraiswamy, P.; Levy, R.; Pikelnaya, O.; Maibach, J.; Feenstra, B.; Polidori, A.; Kiros, F.; Mills, K. Impact of California Fires on Local and Regional Air Quality: The Role of a Low-Cost Sensor Network and Satellite Observations. *GeoHealth* **2018**, *2*, 172–181. [[CrossRef](#)] [[PubMed](#)]
16. Wenger, E. *Communities of Practice: Learning, Meaning, and Identity*; Cambridge University Press: Cambridge, UK, 1999.
17. Gardiner, M.M.; Allee, L.L.; Brown, P.M.; Losey, J.E.; Roy, H.E.; Smyth, R.R. Lessons from lady beetles: Accuracy of monitoring data from US and UK citizen-science programs. *Front. Ecol. Environ.* **2012**, *10*, 471–476. [[CrossRef](#)]
18. Ottinger, G. Buckets of resistance: Standards and the effectiveness of citizen science. *Sci. Technol. Hum. Values* **2010**, *35*, 244–270. [[CrossRef](#)]
19. Simonis, I. Standardized Information Models to Optimize Exchange, Reusability and Comparability of Citizen Science Data. A Specialization Approach. *Int. J. Spat. Data Infrastruct. Res.* **2018**, *13*, 38–47.
20. Kotsev, A.; Schade, S.; Craglia, M.; Gerboles, M.; Spinelle, L.; Signorini, M. Next Generation Air Quality Platform: Openness and Interoperability for the Internet of Things. *Sensors* **2016**, *16*, 403, doi:10.3390/s16030403. [[CrossRef](#)]
21. OGC. Welcome to The Open Geospatial Consortium Available online: <https://www.opengeospatial.org> (accessed on 19 April 2020).
22. Liang, S.; Huang, C.Y.; Khalafbeigi, T. *OGC SensorThings API Part 1: Sensing*, Version 1.0; Open Geospatial Consortium: Wayland, MA, USA, 2016.

23. Tserstou, A.; Jonoski, A.; Popescu, I.; Asumpcao, T.H.; Athanasiou, G.; Kallioras, A.; Nichersu, I. SCENT: Citizen Sourced Data in Support of Environmental Monitoring. In Proceedings of the 2017 21st International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania, 29–31 May 2017; pp. 612–616.
24. Copernicus. Copernicus Land Monitoring Service. Available online: <https://land.copernicus.eu/> (accessed on 19 April 2020).
25. Copernicus. CORINE Land Cover Nomenclature Conversion to Land Cover Classification System. 2010. Available online: https://land.copernicus.eu/eagle/files/eagle-related-projects/pt_clc-conversion-to-fao-lccs3_dec2010 (accessed on 19 April 2020).
26. Nielsen, J. The 90-9-1 Rule for Participation Inequality in Social Media and misc Communities. 2016. Available online: <https://www.nngroup.com/articles/participation-inequality/> (accessed on 19 April 2020).
27. Wikipedia: Statistics. 2020. Available online: <https://en.wikipedia.org/wiki/Wikipedia:Statistics> (accessed on 19 April 2020).
28. van Mierlo, T. The 1% Rule in Four Digital Health Social Networks: An Observational Study. *J. Med. Internet Res.* **2014**, *16*, e33, doi:10.2196/jmir.2966. [CrossRef]
29. Bowser, A.; Hansen, D.; He, Y.; Boston, C.; Reid, M.; Gunnell, L.; Preece, J. Using gamification to inspire new citizen science volunteers. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*; ACM: New York, NY, USA, 2013; pp. 18–25.
30. Iacovides, I.; Jennett, C.; Cornish-Trestrail, C.; Cox, A.L. Do games attract or sustain engagement in citizen science? A study of volunteer motivations. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2013; pp. 1101–1106.
31. Newman, G.; Wiggins, A.; Crall, A.; Graham, E.; Newman, S.; Crowston, K. The future of citizen science: Emerging technologies and shifting paradigms. *Front. Ecol. Environ.* **2012**, *10*, 298–304. [CrossRef]
32. Chou, Y.K. *Actionable Gamification: Beyond Points, Badges, and Leaderboards*; Leanpub: Victoria, BC, Canada, 2015.
33. Fullerton, T. *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*; Taylor & Francis: Abingdon, UK, 2008.
34. Schell, J. *The Art of Game Design: A Book of Lenses*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2008.
35. Dörner, R.; Göbel, S.; Effelsberg, W.; Wiemeyer, J. *Serious Games: Foundations, Concepts and Practice*; Springer: Berlin, Germany, 2016; doi:10.1007/978-3-319-40612-1. [CrossRef]
36. Fuchs, M., F.S.R.P.S.N. *Rethinking Gamification*; Meson Press: Leuphana, Germany, 2014.
37. Xiaomi. Huahuacaocao Flower Care Smart Monitor. 2020. Available online: <https://xiaomi-mi.com/sockets-and-sensors/xiaomi-huahuacaocao-flower-care-smart-monitor/> (accessed on 19 April 2020).
38. Verhoeven, R.; Banasiak, R.; Okruszko, T.; Świątek, D.; Chormanski, J.; Nowakowski, P.; Kardel, I.; Stelmaszczyk, M. Hydraulic Modelling of River Flow—Data collection and problem solving. *Meas. Tech. Data Assess. Wetl. Hydrol.* **2003**, *1*, 9–24.
39. Stone, M.C.; Hotchkiss, R.H. Evaluating velocity measurement techniques in shallow streams. *J. Hydraul. Res.* **2007**, *45*, 752–762. doi:10.1080/00221686.2007.9521813. [CrossRef]
40. Yorke, T.H.; Oberg, K.A. Measuring river velocity and discharge with acoustic Doppler profilers. *Flow Meas. Instrum.* **2002**, *13*, 191–195, doi:10.1016/S0955-5986(02)00051-1. [CrossRef]
41. Smith, R. An Overview of the Tesseract OCR Engine. In Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Parana, Brazil, 23–26 September 2007; Volume 2, pp. 629–633.
42. Vincent, L. Announcing Tesseract OCR. 2006. Available online: <http://googlecode.blogspot.com/2006/08/announcing-tesseract-ocr.html> (accessed on 19 April 2020).
43. Rice, S.V.; Jenkins, F.R.; Nartker, T.A. *The Fourth Annual Test of OCR Accuracy*; Information Science Research Institute: Las Vegas, NV, USA, 1995.
44. Willis, N. Google's Tesseract OCR Engine is a Quantum Leap Forward. 2006. Available online: <https://www.linux.com/news/googles-tesseract-ocr-engine-quantum-leap-forward> (accessed on 19 April 2020).
45. Delta, D. UNESCO World Heritage Centre. Available online: <http://whc.unesco.org/en/list/588> (accessed on 19 April 2020).

46. River, K. Scent Project EU. Available online: <https://scent-project.eu/kifisos-river-basin-attica-greece> (accessed on 19 April 2020).
47. DDNI. Danube Delta National Institute for R&D. 2020. Available online: <https://xiaomi-mi.com/sockets-and-sensors/xiaomi-huahuaacacao-flower-care-smart-monitor/> (accessed on 19 April 2020).
48. SOR. Societatea Ornitologica Romana. 2020. Available online: <https://www.sor.ro/> (accessed on 19 April 2020).
49. HRTA. Hellenic Rescue Team of Attica. 2020. Available online: <http://www.eodathens.gr/> (accessed on 19 April 2020).
50. Forgy, E.W. Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics* **1965**, *21*, 768–769.
51. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [[CrossRef](#)]
52. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 344.
53. Taylor, J.R.; Loescher, H.L. Automated quality control methods for sensor data: A novel observatory approach. *Biogeosciences* **2013**, *10*, 4957. [[CrossRef](#)]
54. Welch, B.L. On the comparison of several mean values: An alternative approach. *Biometrika* **1951**, *38*, 330–336. [[CrossRef](#)]
55. Zar, J. *Biostatistical Analysis*, 5th ed.; Pearson: London, UK, 2010.
56. Krommyda, M.; Theodoropoulos, T.; Sdongos, E.; Amditis, A. Integrated monitoring system for environmental and river data measurements. In Proceedings of the 2019 13th International Conference on Sensing Technology (ICST), Sydney, Australia, 2–4 December 2019; pp. 1–6.
57. DuBois, P.; Foreword By-Widenius, M. *MySQL*; New Riders Publishing: Thousand Oaks, CA, USA, 1999.
58. Ramsey, P.; Columbia, V.B. *Introduction to Postgis*; Refractive Research Inc.: Victoria, BC, Canada, 2005.
59. Böhm, C.; Klump, G.; Kriegel, H.P. XZ-Ordering: A Space-Filling Curve for Objects with Spatial Extension. In *Advances in Spatial Databases*; Güting, R.H., Papadias, D., Lochovsky, F., Eds.; Springer: Berlin, Germany, 1999.
60. Hughes, J.N.; Annex, A.; Eichelberger, C.N.; Fox, A.; Hulbert, A.; Ronquest, M. Geomesa: A distributed architecture for spatio-temporal fusion. In Proceedings of the Geospatial Informatics, Fusion, and Motion Video Analytics V. International Society for Optics and Photonics, Baltimore, MD, USA, 20–24 April 2015.
61. Halldórsson, G.J. *Apache Accumulo for Developers*; Packt Publishing Ltd.: Birmingham, UK, 2013.
62. OGC. OGC Implementing/Compliant Product Details. 2019. Available online: <https://www.openegeospatial.org/resource/products/details/?pid=1554> (accessed on 19 April 2020).
63. Haklay, M. *Citizen Science and Policy: A European Perspective*; Woodrow Wilson International Center for Scholars: Washington, DC, USA, 2015.
64. Kerson, R. Lab for the Environment. *MIT Technol. Rev.* **1989**, *92*, 11–12.
65. Lye, G.C.; Osborne, J.L.; Park, K.J.; Goulson, D. Using citizen science to monitor *Bombus* populations in the UK: Nesting ecology and relative abundance in the urban environment. *J. Insect Conserv.* **2012**, *16*, 697–707. [[CrossRef](#)]
66. Sullivan, B.L.; Wood, C.L.; Iliff, M.J.; Bonney, R.E.; Fink, D.; Kelling, S. eBird: A citizen-based bird observation network in the biological sciences. *Biol. Conserv.* **2009**, *142*, 2282–2292. [[CrossRef](#)]
67. Marshall, N.J.; Kleine, D.A.; Dean, A.J. CoralWatch: Education, monitoring, and sustainability through citizen science. *Front. Ecol. Environ.* **2012**, *10*, 332–334. [[CrossRef](#)]
68. Evans, C.; Abrams, E.; Reitsma, R.; Roux, K.; Salmonsens, L.; Marra, P.P. The Neighborhood Nestwatch Program: Participant outcomes of a citizen-science ecological research project. *Conserv. Biol.* **2005**, *19*, 589–594. [[CrossRef](#)]

