Citizen Science Projects (MOOC) 2.3

Woods, Mel; Coulson, Saskia; Ajates, Raquel; Amditis, Angelos ; Cobley, Andy; Domian, Dahlia

Publication date:
2020

When you are developing a citizen science project, it is crucial to think about what kind of data you want the citizens to collect, from the very beginning. The type of data you collect will influence the shape, format and ways of running of the citizen observatory. It will also determine which citizens are dedicated to your observatory and the technologies they will use. This will be important for the citizens you are trying to engage.

Projects that collect observations are the most common in citizen science. For example, citizens might report observations of phenomena in their own environment. These observations can be collected with or without equipment. For example, you might record each time you see a

particular species of bird in your garden, as in the phenological observatory developed within the Ground Truth 2.0 project [RitmeNatura](http://ritmenatura.cat) or you might use sensors to measure air quality or capture meteorological information, just as the citizens in the GroundTruth 2.0 [MeetMeMechelen observatory](https://mechelen.meetmee.be/) project did, using sensors installed in their bicycles.

Citizens can provide other kinds of data, too. For example, they can validate or interpret scientific data, or contribute to pattern recognition and image analysis, as in the [Adrift project](http://adrift-project.com/), which helps scientists map the ocean trajectories of marine microbes within a simulated web environment.

##Types of data

Most citizen science initiatives are based on people monitoring and registering observations, sometimes using equipment. For this kind of data, the user's identification, the time and place of the observation, the value of the observation and some supporting material like images, audio recordings or videos for validation purposes is usually collected. Observations can be as simple as registering a temperature or as complex as taking a lot of measurements as in the [RiuNet project](http://www.ub.edu/fem/index.php/en/inici-riunet-en), where citizens conduct a complete scientific analysis of several organic and inorganic parameters indicating river water quality.

Observations that use sensor measurements require some additional information, depending on the type of sensor used. You might record the sensor model, date of calibration, sensor parameters, and so on.

These types of observational data include biodiversity observations, environmental monitoring, meteorological observations, hydrological measurements, land cover mapping, and more.

There are other types of projects that are based on the validation of satellite observations with what humans can observe on the ground, the human ability to recognise patterns that computers can't, and the meta-tagging of objects within photos (such as identifying wildlife captured by motion-triggered camera traps. These are all examples of projects that do not collect data directly but instead interpret, analyse or summarise data. [Zooniverse](https://www.zooniverse.org/) is probably the most well-known platform to host these types of projects that citizens can participate in from their homes, no matter their location, just by using a computer. In this way, they can still contribute to the existing data by adding explanations or corrections, descriptions, identifications, and even complete complex data surveys.

A variety of scientific activities can be done in this crowdsourced way, but the data collected in each case will vary widely. The data itself is more important than contextual information like location, time and place, user, or device. The information you want the citizens to provide

should be carefully defined and the surveys carefully designed in consideration of the scientific goal.

##How to decide

To define the type of data you will need, think about time and location coverage of the issue, and the nature of the question you are trying to answer. Questions like ‘where do we need the data to be collected?’ (limited or in a vast zone, predetermined or random, in the field or by a computer) and ‘do we need to repeat observations for the same sample?’, will influence the type of data you collect. These questions will determine what information you require, and also the kinds of technology that you might want to use to capture that information.

Type of information: Analysing the nature of the question will help you to decide which kind of help you need from the citizens. For example, do you need them to check and validate information, sensor variables, or collect samples? These different approaches require different types of data collection. Sometimes, simple text information is enough, but other times, you may be looking for a photograph, sound or video recording, a physical sample or a digital file. Once you know what kind of information you are looking for, you will have a clearer idea of the technology you need to use.

Technology required: Observation projects can be reported on without any sensor or instrument, just using the visual observation of the user, but in most cases, a picture or a record of the observation is desirable, so a camera or other gadget may be needed. Sometimes, you might collect data using sensors, smartphones or computers. If information about the device or sensor will help you understand the data, you may also want to collect that. For example, you can register the kind of instrument, its brand and model, the operating system or its calibrating parameters.

Contextual information: Apart from the data itself, it is normally useful to also have contextual information such as the date and time of collection, location, data collector or conditions.

##Things to keep in mind

Personal data: You will need the consent of your data collectors so that they can participate in activities, especially when you also need to collect some of their own personal data. When you are designing a citizen science initiative, consider whether you need to have your collaborators identified and to what extent. Try to limit the personal data you require from them. Remember that you may be inadvertently collecting personal data that is not relevant to your research question, such as where people live if they are spotting birds in their backyard or the unique identifiers of their mobile phones. Information like this can be used to identify a specific person. It is also essential to clarify what kinds of legal protections or sharing rights that people want to assign to their data when they are contributing photos or measurements. On the

[Natusfera platform](https://natusfera.gbif.es/), for example, each user can decide whether their observations, geolocation and pictures are public or private.

Data quality: The issue of data quality is often raised in the context of citizen science projects, in comparison to official and formal sources of data in scientific research or environmental monitoring. Assuring that the data collected in citizen science projects are of good quality is mostly a matter of keeping processes simple, and making sure that sufficient support, clear instructions, and well-developed protocols are in place. You can also filter the options that citizens can report, or use cross-validation mechanisms where some users validate others' submissions. Documenting the uncertainty of the observations and the dataset overall is a good practice that has been tried in the Ground Truth 2.0 project.

Metadata: Providing adequate metadata both for the observations and for the overall dataset will make it easier to share operations and allow repositories to work together. This also helps scientists to understand the data collected and makes the data more usable. A [Data and Metadata Working Group](https://www.citizenscience.org/get-involved/working-groups/data-and-metadata-working-group/) deals with citizen science metadata within the Citizen Science Association, as does the [Citizen Science Cost Action CA15212](https://www.cs-eu.net/).