22

Maximising the impact and reuse of citizen science data

Jamie Williams[1], Colin Chapman[2], Didier Guy Leibovici[3], Grégoire Loïs[4], Andreas Matheus[5], Alessandro Oggioni[6], Sven Schade[7], Linda See[8] and Paul Pieter Lodewijk van Genuchten[9]

[1] *Environment Systems Ltd, Aberystwyth, Wales UK*
[2] *Welsh Government, Aberystwyth, Wales UK*
[3] *University of Nottingham, UK*
[4] *CESCO, MNHN, Paris, France*
[5] *Secure Dimensions GmbH, Munich, Germany*
[6] *Institute for Electromagnetic Sensing of the Environment, CNR, Milan, Italy*
[7] *European Commission – Joint Research Centre, Ispra, Italy*
[8] *International Institute for Applied Systems Analysis, Laxenburg, Austria*
[9] *GeoCat BV, Bennekom, The Netherlands*

corresponding author email: jamie.williams@envsys.co.uk

In: Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J. & Bonn, A. 2018. *Citizen Science: Innovation in Open Science, Society and Policy*. UCL Press, London. https://doi.org/10.14324/111.9781787352339

Highlights

- Open data and open standards promote interoperability, which in turn allows citizen science data to be more widely discovered and used.
- Data reliability is essential for citizen science data to be trusted and align with environmental regulation and monitoring requirements from governments.
- Contextualising data with metadata, including descriptions of their purpose and methods of dataset creation, allows users to evaluate their possible reuse.
- Reuse of project results is ensured through the use of open data, open standards and by having good data reliability, metadata and documentation.

Introduction

There is an increasing number and diversity of citizen science projects, which can potentially generate new data at a lower cost than professional data collection (De Longueville et al. 2010; Antoniou, Morley & Haklay 2010; Friedland & Choi 2011) and arguably with greater value than those generated by expert knowledge alone (Fischer 2000; and see Danielsen et al. in this volume). When considering Ten Principles of Citizen Science (ECSA 2015), in particular openness and accessibility, these citizen science data have the potential to be a valuable source of information for decision-making and policy formation on local, regional and national scales. However, for the data to realise their full potential, a number of factors have to be considered.

This chapter identifies the factors that affect citizen science data using examples from environmental monitoring and geographic information. These factors include open data standards and interoperability; data reliability and alignment with government environmental regulation and monitoring requirements; the contextualisation of data to enable users to evaluate its possible reuse; and the reuse of project results (see box 22.1). This chapter addresses each of these factors in turn to help specialists and non-specialists alike to better plan citizen science projects.

Data contextualisation

Data are only meaningful if they can be interpreted. Therefore, it is vital to know the context within which a particular dataset has been created,

Box 22.1. Key factors when initiating a citizen science project

- **Data contextualisation –** communicating the context in which a particular dataset has been created.
- **Data interoperability –** enabling seamless reuse of resources (in this case, data and processing) across different systems.
- **Data quality –** data quality has long been identified as the crucial challenge for the use of citizen science data.
- **Data reuse –** data ownership and future accessibility.

including, for example, units of measure, measurement devices, pre-processing procedures, quality assurance (QA) mechanisms, uncertainties and intended use. As soon as a dataset needs to be understood by anybody other than by its creator (for example, by a customer, reviewer or peer), then this contextual information should be explicitly provided, ideally with the dataset itself.

Some of the contextual description of why, how, when and by whom a dataset was created can be unambiguously provided using standard vocabularies and code lists (for example, for units of measure or particular statistical processing algorithms). However, other information may bring uncertainty when interpretation takes place. The description of data provenance (i.e., the processing steps applied) is a typical case in which it remains difficult to keep track of a complete, up-to-date and reproducible instruction. This holds particularly for datasets subject to intense experimentation. In citizen science, these issues are complicated by the training required to provide detailed contextual information about data, which may not be readily available to participants.

Context again becomes relevant if a given dataset is applied to another, initially unintended, purpose. In such cases, the effects of the contextual change on possible interpretations of the data have to be carefully examined. For example, in establishing if occurrences of fly fishing derived from social media can be used for an indication of river health.

The creation of metadata is key to capturing contextual information. It would normally consist of the title, description, number of participants/observations, contact details, and temporal and geographical extent of the data. If the data are combined with other data, legal constraints and data quality aspects such as lineage information should also be included. For the purpose of assessing data quality, the identity of the observer, location accuracy and potentially the device accuracy would also be required.

Citizen science projects are constantly looking for new participants, and ways to make their project's data available through as many means as possible, therefore making the efforts of the voluntary work as effective as possible. Wide discovery of citizen science resources is important for maximising impact, creating additional value and encouraging reuse, beyond the scope of the original project.

In support of this, scientists, journalists and citizens are continuously looking for datasets relevant to their field of study, typically querying a search engine or open data catalogue for datasets using keywords and a location/timeframe of interest. The use of common vocabularies makes metadata meaningful for different usages, for example, DCAT enables interoperability with the open (government) data, schema.org enables discoverability and ISO19139 connects to the spatial data infrastructure (SDI) community.

Data interoperability

As this book illustrates, citizen science can be extremely varied, with diverging objectives and research questions from within and across different subject areas and geographic scales. The level and type of citizen participation also varies greatly (see Haklay; Ballard, Phillips & Robinson, both in this volume). Consequently, data are generated, analysed and presented in a variety of ways. The case-by-case, tailor-made management and handling of datasets might serve their original intended purpose (for example, assessing water quality or monitoring birds) but is likely to reduce interoperability, in other words, the communication, exchange and use of data. This not only includes the interaction between machines, but also between machines and humans (users), and between communities of people themselves.

Interoperability, which enables the seamless reuse of resources (in this case, data and processing) across different systems, can be reached by applying community-wide agreements. For example, agreeing on the use of software tools, data standards and best practices, or by improving data accessibility and exchange through ad hoc tools or community practices. It can address the data themselves, but also the processes and services that generate and exchange data between any two parties (for example, between citizens and academics working on a national-funded research project, and between citizens and local decision-makers collaborating for the benefit of the local community). Such processes help to i) ease the integration of data from different sources; ii) improve the reuse of data in other contexts; and iii) save resources in the development of data management and handling tools. Semantic interoperability provides interoperability at the highest level to exchange data with unambiguous, shared meaning and improve quality, efficiency and efficacy. Semantic interoperability adds to the possibilities of data sharing as well as the meaning of the data, linking any data elements with metadata and terms of vocabulary.

There is a high diversity in the details of modelling, encoding and describing datasets, as well as in the communication protocols for data storage, processing and access. Museum collections, institutional biodiversity datasets and international projects have adopted extensions of the Darwin Core in multiple encodings, while the SDI favour other markup languages for sharing geographical data and metadata. Initiatives such as the PPSR_CORE Program Data Model Metadata Standard/data-sharing protocol are progressing this field. However, each initiative has a form of standardisation in mind, which prevents each new citizen science project from designing their own ontology relating to their domain. Once a known subset of standards for specific domains has been developed, it will be much easier to write connectors (mappings) to interact between those standards, therefore greatly improving the data-sharing potential.

Data standards (i.e., user-accepted norms on data models, formats and exchange protocols) are the key to achieving interoperability. The main challenges include the agreement of standards within a user group and across communities. The Open Geospatial Consortium (OGC), for example, has a long history in the technical specification of geographic

Box 22.2. EC INSPIRE Directive

'The INSPIRE Directive aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment' (European Commission 2017). It is based upon these common principles:

- "Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used".

data and metadata (i.e., data documentation) models and encodings. The Sensor Web Enablement for Citizen Science (SWE4CS) proposal for a standard in the OGC exemplifies the need for flexibility when modelling and exchanging citizen science data (OGC/CS DWG 2016). In some regions, standards are complemented by more conceptual frameworks, such as the legally binding European Directive 2007/2/EC to establish an Infrastructure for Spatial Information in the European Community (INSPIRE) (European Commission 2017; see box 22.2), and the US Crowdsourcing and Citizen Science Act of 2016 (Congress 2016).

Experts are now widening the extension of these standards into the citizen science community. Currently, SWE4CS (Citizen Observatories 2015), exists in parallel to a Volunteered Geographic Information (VGI) extension to the INSPIRE standards (Reznik et al. 2016). SWE4CS also enables the inclusion of concepts from other standards that are established in other areas. See box 22.3 for examples of common metadata standards for dataset discovery.

As previously mentioned, metadata plays an important role in the data-sharing process. Beyond just providing an understanding of what is published, machine-to-machine understanding needs standardised interfaces with common exchange formats. This is not a requirement for citizen science data but would maximise their use and reuse. Few citizen science projects currently adopt standards for web services or data encodings, as most projects have yet to realise the benefits of sharing their data or are unaware of how best to do so. Agreements and technology implementations are needed, first to adapt established practices to new interoperable systems, and second to stimulate new projects to adopt standards and tools.

Ongoing global initiatives such as the Group on Earth Observations Biodiversity Observation Network (GEO BON), which 'aims to improve the acquisition, co-ordination and delivery of biodiversity observations and related services to users including decision-makers and the scientific community' (GEO BON 2017), and Global Biodiversity Information Facility (GBIF) 'an open-data research infrastructure funded by the world's governments and aimed at providing anyone, anywhere access to data about all types of life on Earth' both provide guidance on aspects of interoperability within environmental monitoring (GBIF 2017). However, more needs to be done to get novice, local-scale citizen science projects to adopt these standards.

Box 22.3. Examples of common metadata standards for dataset discovery

Dublin Core	Vocabulary for resource description, used as a base vocabulary in other vocabularies (http://dublincore.org/documents/dces/). DC is the default schema in Catalogue Service for the web, a metadata transfer protocol standard by Open Geospatial Consortium (http://www.opengeospatial.org/standards/cat)
ISO19139	XML/XSD-based vocabulary to describe spatial datasets (https://www.iso.org/standard/32557.html), commonly used in the GIS domain (INSPIRE).
DCAT	Data Catalog Vocabulary is a Resource Description Framework vocabulary to describe datasets maintained by W3C. Used in open data portals (http://www.w3.org/TR/vocab-dcat/).
VOID	Vocabulary of Interlinked Datasets is a vocabulary to describe linked datasets maintained by W3C (https://www.w3.org/TR/void/).
schema.org	Initiative of the main search engines to enable crawling web content as structured data. Contains a concept for dataset (http://schema.org/Dataset).
SDMX	Vocabulary to describe datasets in the statistical domain (https://sdmx.org/).
Datapackage	Vocabulary to describe (and embed) datasets, maintained by Open Knowledge Foundation (https://specs.frictionlessdata.io/data-package/).

Data quality

It is commonly agreed that the lack of knowledge about data quality limits the use of citizen science data (Flanagin & Metzger 2008; Haklay 2010; Goodchild & Li 2012; Fowler et al. 2013; Hunter, Alabri & Ingen 2013). Furthermore, citizen science projects are designed to be carried out by non-experts, with controlled data collection methods to support scientific integrity (Craglia & Shanley 2015). With project goals of i) enlarging participation and consequently data collection over space and time, and ii) ensuring that information embedded in the dataset varies according to the field of interest, context such as location, date and rules for data standardisation should be specified as part of the project, but parameters for participation, such as skill level, interpretation and observation intensity, should remain flexible. However, where data collection protocols are not respected by participants or simply implemented incorrectly, the resulting data could be of lower quality and potentially include misleading information. To further confound this, a learning effect has been reported in many citizen science studies, for example, participants get better at identifying different species (see box 22.4; Peltola & Arpin in this volume), though their ability may differ between activities. To balance this, the iSpot crowdsourcing qualifying system, for example, uses a reputation score for participants over eight groups of species. The contributor's reputation per species group acts as a quality measure of trust and can be used to evaluate their identifications over alternatives. Using this system, Silvertown et al. (2015) reported improvements in accuracy when multiple identifications were recorded, as well as the ability to quantify the level of confidence in observations.

Some data quality issues can be addressed by using a QA process, either human or automatic, to produce metadata on data quality. This quality information establishes trust in an observation and the volunteer who produced it, in a similar way to the trust traditionally placed in experts (Alabri & Hunter 2010; Hunter et al. 2013; Bishr & Kuhn 2013; Zhao et al., 'A Spatio–Temporal VGI Model', 2016). This trust is then transferable to the data themselves (Leibovici et al. 2017a).

A human-based QA process, such as peer verification, allows project participants to help identify and validate the observations provided by new users (Warncke-Wang et al. 2015; Antoniou & Skopeliti 2015). This peer verification, crowdsourcing the quality assessment or 'wisdom of the crowd' (Surowiecki 2005), enables some control in the same way that Wikipedia allows editing of an article to support convergence towards shared narratives. In Wikipedia, the data themselves are subject to peer verification quality improvements and the edits are logged (Warncke-Wang et al. 2015; Mobasheri et al. 2015). For citizen science, most of the time the data will not be as modifiable as in Wikipedia. However, they will have a quality that may be identifiable, and according to the level of quality and reliability that was attributed to the data, it may be reused regardless of whether it has been validated or not. Nonetheless, multiple citizen science observations in the same location or made at the same time can allow for application of a similar process as in Wikipedia editing. Volunteered geographic information (VGI) such as OpenStreetMap (OSM) data follows this principle (Haklay 2010).

Peer verification, such as expert verification, is not without issues (Wiggins et al. 2011; See et al. 2013). The volume of data to be checked and verified can be overwhelming, and errors made in human verifying may still have implications. The development of geo-computational QA offers greater scalability with constant reliability of assessment, ensuring better comparability (Kelling et al. 2011; August et al. 2015; Meek, Jackson & Leibovici 2016; Leibovici et al. 2017a). With this, stakeholders setting up a citizen science project can define the QA with its requirements based on rules defining the levels of quality, which are then transformed into a workflow of quality controls, generating the metadata on data quality.

It is possible to use automatic QA to complement peer assessment, accumulating trustworthiness in the volunteers (Leibovici et al. 2017b) (see box 22.4). The impact of these different QA methods can vary, so as part of the whole data curation process the QA has to be designed, agreed to for the usage of the citizen science data, and published as part of the metadata (Higgins et al. 2016).

Assessment standards are still to be finalised to communicate metadata on data quality, and will be an addition to the interoperability discussed above. Currently, the ISO19157 metadata standard for geographical data is applicable to citizen science as it produces geolocated data. This includes 'usability' but omits quality dimensions, such as trust and number of participants. Therefore, on top of the 'producer' model represented by ISO19157, citizen science demands a 'stakeholder' model to assess the participant (as a sensor) and a 'consumer' model through which peers give feedback on an observation (Meek, Jackson & Leibovici 2014; Leibovici et al. 2017a).

Data reuse

Supporting and planning for the reuse of data collected through citizen science activities is key for realising their long-term value. There are several aspects that need to be considered when planning for sustainable data management, such as the intellectual property rights (IPR) associated with contributions from citizens with respect to patents and copyrights (Scassa & Chung 2015a). The raw data contributed to a citizen science

Box 22.4. Quality assurance in an invasive species survey of *Fallopia japonica* (Japanese knotweed) in Wales (Leibovici et al. 2017b).

This example is typical of a plant identification survey and illustrates the different dimensions of quality that are important in citizen science. Participants were trained to identify Japanese knotweed and were sent to the Snowdonia National Park in Wales to locate, capture geolocated pictures and answer questions on invasive species. Reliability in the location and identification of the plant were the most important quality assessment criteria.

When it is not possible to manually assess each observation for accuracy, rules can be established to help assess the observation based upon factors such as proximity to cultivated land and forest, rivers or paths, but these can in turn be compromised with poor positioning (propagation of error). Modelled Earth observation data can also be used to assess the likelihood of species presence although this includes problems such as satellite imagery from different dates to that of volunteer data capture. Confirmation from multiple observations or closeness of observations can also be used.

Combining these imperfect rules and quality controls could lead to an improved QA. Furthermore, adding bespoke rules concerning the interaction of the factors may also improve the final assessment, and therefore their reusability.

project have no copyright, but the form in which the data are presented may qualify for copyright, for example, with photographs or written text. Therefore, it is important for citizen science projects to consider what contributions might be subject to IPR and the form in which the contributions are made.

To help developers of citizen science projects consider these issues, Scassa and Chung (2015a) provide a typology that categorises citizen science projects into four types: i) classification or transcription of data; ii) data collection; iii) participation as a research subject; and (iv) problem-solving, data analysis or development of ideas. They argue that there will be minimal IPR issues related to the first three categories based on examinations of the form of participation in different citizen science projects, for example, those found on Zooniverse. However, they also identified examples of projects that collect photographs or written text and therefore may be subject to copyright issues. The fourth type of citizen science project has potential patent issues since citizens may engage in inventive activities that could lead to patent rights, of which developers of citizen science projects should be aware. Within the EU there are additional database rights provided through the EU Database Directive. Licensing is one way to handle these IPR issues. For example, OSM has an Open Database License that specifies use of the data by anyone for any purpose provided attribution is given to the project and its contributors as a whole.

When initiating a citizen science project, the two primary high-level considerations regarding IPR and citizen science are (1) what background IPR will be used (for example, knowledge and data) and what restrictions is it subject to; and (2) if the project wishes to allow access to the knowledge and data (and to what level – see box 22.5) generated by the project (foreground IPR). Guidance on IPR is available from multiple governing bodies, organisations and institutions. Ultimately, how the IPR for any citizen science project is handled should be set out in the terms of participation in a project (the 'terms of use') (Scassa & Chung 2015a) so that participants are clear on these conditions and can agree to them during registration, prior to data collection.

A further consideration for data reusability is personal privacy. Protecting participating citizens' privacy is a key priority in a citizen science project (Bowser et al. 2014). When individuals provide data as part of a citizen science project, data may be stored with the individual's personal details. The contribution may sometimes require that the citizen's identity is known or can be known if necessary, or the citizen might wish to restrict, or actively promote, the attribution of their contribution with their personal details. Location-based information, recorded using mobile devices, can further reveal the position of individuals as well as their movements. This information could inadvertently be used to locate individuals in space and time, and in some cases, identify their home address or workplace. Regardless of the intended use of the personal information collected during and after the project, it must be stated in projects' 'terms of use' at the point of registration or, as a minimum, prior to commencement of data collection.

Many countries have data protection laws that protect individuals; however, these vary from country to country (Dyson et al. 2014). For Europe, the EU Regulation 2016/679 will protect EU citizens in terms of the processing and free movement of personal data. This regulation comes into force in 2018. Some principles are that users must be able to control their personal data at any time, including the inspection and deletion of

Box 22.5. Open data

The proliferation of open data can bring new opportunities in environmental monitoring (and other areas), by allowing cross-validation, data conflation, or increased temporal or spatial coverage. Citizen science data have a role to play in this open data movement, where open data is defined as the following:

> Open data is publicly available data that can be universally and readily accessed, used, and redistributed free of charge. Open data is released in ways that protect private, personal, or proprietary information. It is structured for usability and computability. (Verhulst & Young 2016)

One way in which citizen science data can be released as open data is using the Creative Commons (CC) open data licensing framework. Creative Commons encourages sharing under any of its licences as a way to create a more open data culture. Examples of CC licences that conform to the open definition and which could be suitable for use with citizen science applications (providing this intent was stated in the project's terms of participation, prior to citizen participation) are:

- Creative Commons (CC0),
- Creative Commons Attribution 4.0 (CC-BY-4.0),
- Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0).

By releasing project data under one of the established CC licences, thus allowing any restrictions on their use to be fully understood, the likelihood of data reuse greatly increases.

their personal record. Personal data can only be collected for a particular purpose and the user must agree to this (i.e., give prior consent) before the data are exchanged. The personal data collected should also be limited to what is absolutely necessary, which requires knowing this information for a given project in advance. This regulation means that implementation of pan-European citizen science platforms can be challenging. Moreover, this becomes problematic if there is personal data exchange to countries outside of the EU, where there is none, or a variation in personal data protection legislation. One example would be the

United States, which has passed the Crowdsourcing and Citizen Science Act of 2016. This act endeavours, 'While not neglecting security and privacy protections', to make data collected through a citizen science project open and available, in machine-readable formats, to the public. As part of this process, federal agencies are required to inform participants on the expected uses of a project's data and if project results will be made available to the public. Furthermore, federal agencies would retain ownership of such data.

Data may also be collected about people, species or other entities that exist in real life. In this context, privacy and security not only play a major role to ensure the protection of personal data from citizens but also the well-being of the objects observed. Two types of 'objects' can be identified: i) primary objects about which the citizen is collecting information, for example, the ancient tree in a photograph; and ii) secondary objects that are recorded with the primary object, either by accident or because it was not possible to record just the primary object, for example, if the ancient tree was located alongside a school with children playing outside, who are also captured in the photograph. The spectrum for protection of these objects is manifold but can be condensed to conditions that apply when making observations electronically available. For example, the observations in time and location for endangered species are one type of information that may not be made available to the general public (primary

Box 22.6. Data contextualisation, interoperability, quality and reuse, in practice

An example of the key considerations for citizen science projects in practice is Geo-Wiki, which is an online platform and set of mobile tools for improving global land cover datasets (Fritz et al. 2012). Geo-Wiki was designed to address the problem of the high spatial disagreement that can be observed when different global land cover products are compared (Fritz et al. 2011) and to use data collected by citizens to create improved hybrid land-cover maps (See Fritz et al. 2015). In the online tool, citizens are asked to interpret land cover using medium-resolution satellite imagery from Google Earth and Bing; and more recently images from the Sentinel-2 satellite have been added. In the mobile apps, citizens are guided to locations and asked to classify the surrounding land cover and land use, supplementing their observations with geo-tagged photographs. More

(continued)

opportunistic tools are also available for recording land cover and land use at any location.

To involve citizens in the data collection process, the Geo-Wiki team have run a number of citizen science campaigns that have lasted a few weeks to six months. Various incentives are used to encourage participation, from prizes to co-authorship on scientific papers. More details of various campaigns can be found in See et al. (2015); Sturn et al. (2015); and Laso Bayas et al. (2017).

The four key issues discussed in this chapter have been tackled by the Geo-Wiki project. The raw data collected during the first set of citizen science campaigns have now been published in an open-access repository, PANGAEA (Fritz et al. 2017). Data from a more recent campaign focusing on classification of imagery for crop-land has also been published in PANGAEA (Laso Bayas et al. 2017). This publication of the data supports both the **interoperability** and **reuse** of the project by encouraging reuse of the data for applications such as land cover map development (as training data) and for the evaluation of land cover maps (i.e., validation). Although the data do not follow a specific metadata standard, they are supplied with accompanying metadata that explains each of the data fields. The **contextualisation** is provided through the narrative that accompanies the publication of the data (Fritz et al. 2017; Laso Bayas et al. 2017) and the land-cover definitions used are generic enough that they can be applied to many other land-cover products.

The data have been published open access in raw form so that users can apply their own data quality measures to the observations and filter them based upon the needs of their own applications. The **data quality** has been analysed and reported in a number of different papers using a variety of methods, ranging from comparison with authoritative or expert data sources to different conflation methods such as majority voting when multiple observations are available for a single location (See et al. 2013; Laso Bayas et al. 2017; Salk et al. 'Local Knowledge', 2016; Salk et al., 'Assessing Quality', 2016; Salk et al. 2017; Zhao et al. 2017). Overall the reliability has been good and lessons have also been turned into recommendations to further improve it, such as methods to enhance the training of the citizens, more effective use of real-time feedback, and so forth.

objects). In general, observations that contain children or sexual content as well as violent language must be redacted according to the Western law. Furthermore, personal identifiers such as car license plates, doorbell signs and people who did not agree to be visible (secondary objects) must also be removed. This ensures the privacy and well-being of the secondary objects. Automatic or semi-automatic processing of this typically involves functions that can be identified as a privacy/security extension to QA, where the objective is to be compliant with the governing law.

Conclusions

Citizen science data can act as timely evidence for various decision-making processes that impact on citizens' lives and surroundings, including environmental policy. However, it is only with good management of data and metadata, particularly when it comes to data reliability, that citizen science data can fulfil their role of empowering citizens.

Establishing the evidence that citizen science can be used effectively for policy will take time (see Nascimento et al.; Shirk & Bonney, both in this volume). However, in order for policy to realise its full benefit, the ability to share and use data across platforms and stakeholder groups is essential (Higgins et al. 2016). To maximise the impact and reusability of citizen science data, citizen science projects should therefore adopt standards for web services or data encodings and, where possible, adapt previously collected observations to these standards. This allows other citizen science initiatives on the same or complementary topic to reuse the data generated.

Being able to ingest, conflate and disseminate citizen data across systems not only supports the generation, assessment and sharing of citizen science data as evidence suitable for decision-making, but also improves its impact and reusability. This is achieved through the transmission of data to other interoperable systems, used by other projects and purposes, to stimulate more targeted research, societal benefits and potentially commercial revenue.

The field of citizen science is both long-established and continually evolving. New technologies and understandings provide the potential to increase the impact of citizen science projects. The Ten Principles of Citizen Science offer some guidance in the area of maximising the impact and reuse of citizen science data (ECSA 2015). However, further technical work is needed in the areas of domain-specific citizen science metadata and data quality (among others). Additionally, specific guidance on the areas of IPR and privacy would contribute towards citizen science data reaching their full potential. In support of this, dedicated organisations such as the European Citizen Science Association (ECSA), the Australian Citizen Science Association (ACSA) and the Citizen Science Association (CSA), and other organisations with domain-specific working groups, such as the OGC and the Committee on Data for Science and Technology (CODATA), among many others, continually work towards these common goals.