International Journal of Digital Earth

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tjde20

Assessing FAIRness of citizen science data in the context of the Green Deal Data Space

Victoria Lush, L. Bastin, K. Otsu & J. Masó

To cite this article: Victoria Lush, L. Bastin, K. Otsu & J. Masó (2024) Assessing FAIRness of citizen science data in the context of the Green Deal Data Space, International Journal of Digital Earth, 17:1, 2344587, DOI: 10.1080/17538947.2024.2344587

To link to this article: https://doi.org/10.1080/17538947.2024.2344587

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2024, VOL. 17, NO. 1, 2344587
https://doi.org/10.1080/17538947.2024.2344587

REVIEW ARTICLE

OPEN ACCESS Check for updates

Assessing FAIRness of citizen science data in the context of the Green Deal Data Space

Victoria Lush [a], L. Bastin [a], K. Otsu [b] and J. Masó [b]

[a]Computer Science, Aston University, Birmingham, UK; [b]Grumets Research Group, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain

ABSTRACT

As part of the European Data Strategy, the European Commission is working on common European data spaces, including a Green Deal Data Space (GDDS) that covers issues such as climate change, circular economy, pollution, biodiversity, and deforestation. The successful development of the EU GDDS will depend on the availability of FAIR (findable, accessible, interoperable, and reusable) data sources, including FAIR citizen science data. While the importance of FAIR principles is increasingly acknowledged within the field of citizen science, sources of FAIR data outside the biodiversity domain are generally scarce. This is contributed by the lack of end-to-end technical solutions, readily available semantic resources to support data interoperability, and centralised data repositories suited for citizen science data. To investigate the current state of play with citizen science data FAIR compliance, we conducted a review to elicit platforms, tools and standards either used by or indicated as suitable for facilitating stages of the citizen science project lifecycle. We report on the results of our review and discuss gaps that still exist to achieve citizen science data FAIRness. We also examine three data aggregation platforms identified in our review which closely align with FAIR, namely: the Global Biodiversity Information Facility, OpenStreetMap, and Sensor.Community.

ARTICLE HISTORY
Received 22 September 2023
Accepted 14 April 2024

KEYWORDS
Citizen science; FAIR guiding principles; open data; Green Deal Data Space

1. Introduction

To overcome the challenges of climate change and environmental degradation, the European Union has adopted the European Green Deal (EC 2020) as a way '*to transform the EU into a modern, resource-efficient and competitive economy, ensuring: no net emissions of greenhouse gases by 2050; economic growth decoupled from resource use; no person and no place left behind*' (EC 2023).

As part of the European Data Strategy towards establishing a Single EU Market for data, the European Commission proposes European data spaces across several domains to enable easy data flow between countries and sectors (EC 2022). This includes a Green Deal Data Space (GDDS) covering issues such as climate change, circular economy, pollution, biodiversity, and deforestation (Farrell et al. 2023). To be useful to decision makers, the GDDS will need FAIR (findable, accessible, interoperable and reusable) data sources (INSPIRE 2022).

- The '*FAIR Guiding Principles for scientific data management and stewardship*' were published in 2016 to enhance the value of digital assets (Wilkinson et al. 2016). These principles specifically

CONTACT Victoria Lush lushv1@aston.ac.uk Computer Science, Aston University, Aston Triangle, Birmingham, UK

focus on machine capability to automatically find, access, interoperate with and reuse assets, thereby promoting open science, which may be defined as: '*a collaborative culture enabled by technology that empowers the open sharing of data, information, and knowledge within the scientific community and the wider public to accelerate scientific research and understanding*' (Ramachandran, Bugbee, and Murphy 2021).

While there is no full agreement in the scientific community on how FAIR should be evaluated in practice (Peng 2023), the Go FAIR initiative defines a FAIR assessment framework consisting of 10 principles and sub-principles, with a total of 15 criteria (GO FAIR 2016).

1.1. Citizen science

The term 'citizen science' primarily emerged from the field of biodiversity (Bonney et al. 2009); it can be defined as general public or non-expert participation in scientific processes to produce or enrich scientific knowledge (Eitzel et al. 2017). There is no single agreed definition of citizen science; a comprehensive list of definitions can be found in Haklay et al. (2021).

While the main purpose of citizen science projects varies, most collect data as a part of their activities. This data can play an important role in complementing official data sources (Fritz, Costa Fonte, and See 2017; Haklay, Mazumdar, and Wardlaw 2018; König et al. 2021; Sullivan et al. 2014; Sy et al. 2020), not least because of its currency and specificity (e.g. Ferri et al. 2020). While quality of such data remains a concern (Aceves-Bueno et al. 2017; Stevenson, Merrill, and Burn 2021; See 2019), the development of consistent study protocols, advanced data collection and data visualisation tools, data standards and protocols, machine learning for calibration and outlier detection can help address some data quality issues (Balázs et al. 2021; Fraisl et al. 2022; See 2019).

Active citizen engagement in science is now one of the European Research Area priority actions, as defined in the Pact for Research and Innovation (R&I) in Europe (Council of the EU 2021). The Open Science Policy of the European Commission recognises citizen science as one of its eight policy ambitions, stating that '*the general public should be able to make significant contributions and be recognised as valid European science knowledge producers*' (EC 2019). These developments underline the importance of including citizen science data within the GDDS. However, to be fit for integration into any Data Space or be effectively and properly re-used outside of the project that collected the data, citizen science data needs to adhere to the FAIR data principles. This is supported by the '*10 Principles of Citizen Science*' developed by the European Citizen Science Association (ECSA) where Guideline 7 states that '*Citizen science project data and meta-data are made publicly available and where possible, results are published in an open access format*' (ECSA 2015), implying adherence to FAIR principles.

The goal of this work is to investigate the current state of play regarding FAIRness of citizen science data. As such, we explore the following research questions:

(1) What platforms, tools and standards are currently used by the citizen science community to collect, document and share citizen science data?
(2) Can the tools used by the citizen science community effectively support the production and governance of FAIR data?
(3) What gaps still exist to support FAIR citizen science data?

To address these research questions, we conducted a review of scientific publications, citizen science conference publications and major citizen science platforms to identify tools, platforms, standards, and standardised resources used by, or suitable for running citizen science projects. In this paper, we present the results of our review and examine three longstanding initiatives and data aggregation platforms that successfully apply open standards and tools to collect and share community-generated data and are relevant to the GDDS.

In Section 2, we outline the review method, while Section 3 discusses major citizen science project discovery platforms. Section 4 examines three initiatives which closely align with at least some of the FAIR principles. Sections 5 discusses the tools that can support citizen science projects at different lifecycle stages and considers open standards that could support citizen science data FAIRness. Finally, Section 6 presents conclusions and discusses gaps that still exist for enabling FAIR citizen science data.

2. Method

The search for scientific papers was conducted in March 2024 in Scopus and Web of Science by searching ('Citizen Science' AND FAIR) keywords in 'title', 'abstract', and 'author keywords'. The search returned 123 results (plus one collection of 28 papers related to FAIR in citizen science). After removing 41 duplicates and screening for relevance, the final set consisted of 32 publications. Only those that directly focused on citizen science and discussed FAIR principles were considered.

We additionally conducted a review of (1) ECSA 2022 Conference Proceedings, (2) S*Csi 2023 Conference abstracts and poster presentations, and (3) information resources available on the EU-Citizen.Science, SciStarter, CitizenScience.gov, CSA, ECSA platforms. This was done to capture current trends in citizen science project management since many projects may not have resources, sufficient scientific results, or awareness to publish their work in peer-reviewed conferences and journals.

Each source was reviewed to elicit platforms, tools and standards either used by or indicated as suitable for facilitating stages of the citizen science project lifecycle: project hosting, data collection, data documentation, data storage, and data publication and sharing. Figure 1 presents the identified platforms, tools and standards. Table 1 lists discovered information resources relevant to open data and FAIR principles. Table 2 lists semantic resources identified in the review; these relate to the interoperability facet of FAIR. Detailed summaries of the results available in Appendix 1 and 2.

3. Citizen science project discovery platforms

Citizen science project discovery platforms facilitate the search of citizen science projects hosted on the platform itself and/or other sites. Such platforms might be considered as an obvious choice to search for data; however, at present, these primarily focus on project discovery by prospective participants or collaborators and the provision of guidelines and resources, rather than on the curation of project data. Our review identified three project discovery platforms. Here, we focus on EU-Citizen.Science and SciStarter since CitizenScience.gov only supports the US federal government projects.

EU-Citizen.Science[1] was established in 2019, initially funded by the EU Horizon 2020 programme, and now supported by a consortium of 14 partners and nine third parties. It primarily focusses on projects within the EU but is not exclusive to Europe. It contains 271 projects and, in addition to project discovery, offers 220 information resources, a Moodle Training Platform with 24 training courses, and a Swagger API for retrieving full project metadata.

SciStarter[2] was founded in 2011 and is primarily supported by grants from the National Science Foundation, Institute for Museum and Library Services, Schmidt Futures, NASA, and National Library of Medicine. SciStarter is a global platform covering a range of thematic areas and is more popular among US-based projects. It contains 1528 registered projects, and 426 free and low-cost tools (e.g. designs for sensors and testing kits) for making observations, recording data, and processing samples. The platform offers data hosting, which allows users to submit their observations and permits the visualisation of observations on a map. This enables potential re-users to better evaluate project data for fitness-for-use; however, raw observation data is not accessible to download.

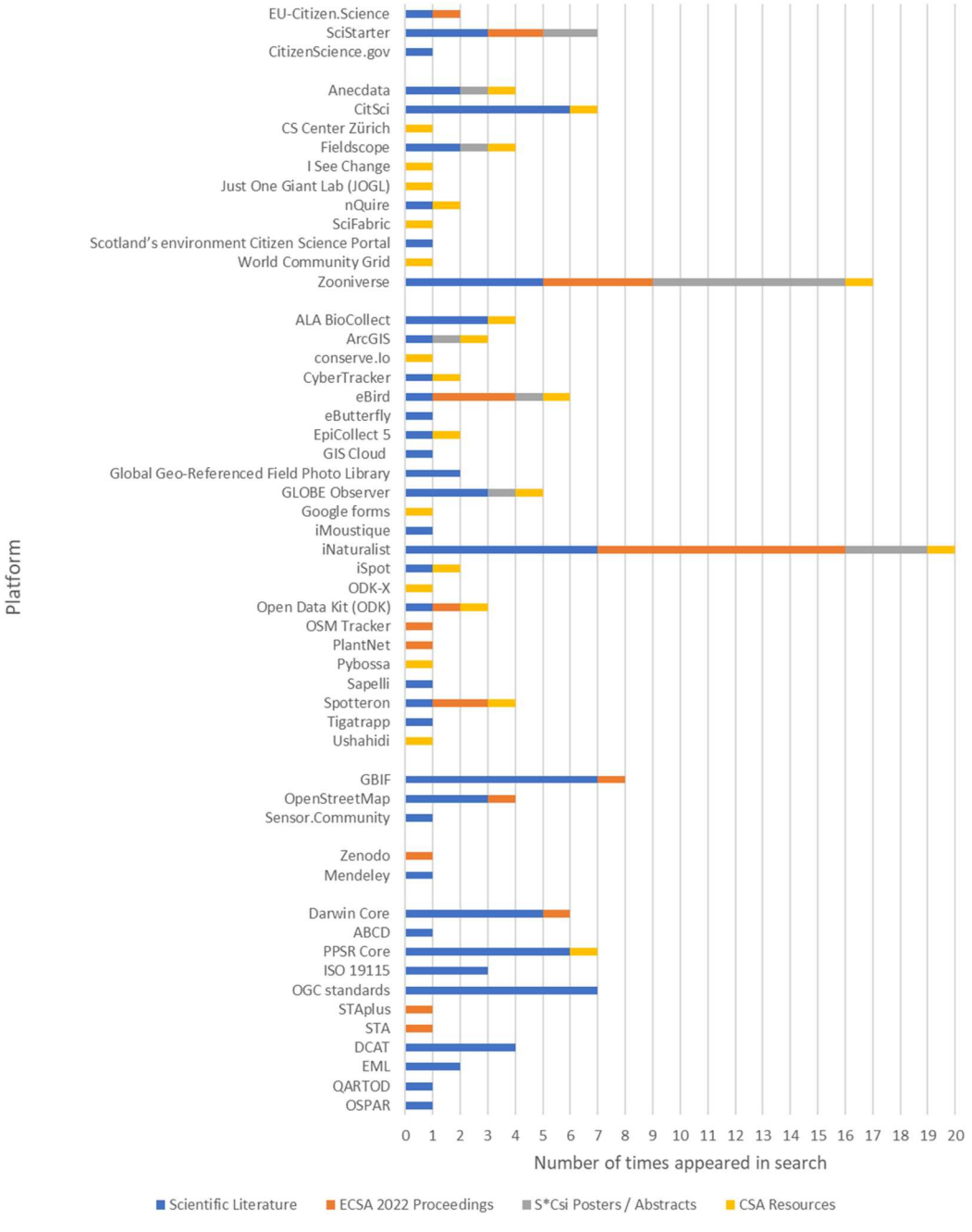

Figure 1. Tools, platforms and standards identified in the review.

Table 1. Information resources related to FAIR and open data identified in the review.

Resource	Source
FAIR Data in Citizen Science Projects	EU-Citizen.Science
Doing Citizen Science as Open Science	EU-Citizen.Science
Basic Regulations and Ethics for Citizen Science	EU-Citizen.Science
UK Environmental Observation Network (UKEOF) Resources	EU-Citizen.Science / UKEOF
Data Ethics for Practitioners	SciStarter

Table 2. Semantic resources identified in the review.

Resource	Source
AGROVOC	Coché et al. 2021
Chemical Entities of Biological Interest (ChEBI)	Ramírez-Andreotta et al. 2021
Darwin Core	Turicchia et al. 2021; Coché et al. 2021
Environment Ontology (ENVO)	Ramírez-Andreotta et al. 2021
GEMET/INSPIRE	Coché et al. 2021
NERC Vocabulary Server	Turicchia et al. 2021b
The Exposure Ontology	Ramírez-Andreotta et al. 2021
University of Arizona Superfund Research Project Data Interface Ontology (SRPDIO)	Ramírez-Andreotta et al. 2021

Project discovery platforms deliver the vital function of promoting citizen science projects and offering resources and training for citizen science practitioners. While such platforms continue to grow and evolve (as demonstrated by the large number of projects and resources listed on EU-Citizen.Science and SciStarter), these are unlikely to serve as centralised citizen science data hubs due to the lack of necessary technical resources and data licensing issues, as there is no obligation for the projects to provide open data.

4. Data aggregation platforms

Our review identified three large-scale initiatives and data aggregation platforms which closely align with at least some of the FAIR principles and are relevant to the GDDS: The Global Biodiversity Information Facility (GBIF), OpenStreetMap and Sensor.Community. We examine these initiatives highlighting their approaches to achieving data FAIRness.

Before diving into the discussion, it is important to note two structures of governance of public participatory science, namely ***top-down*** and ***bottom-up*** approaches.

The **top-down** approach traditionally refers to the type of governance where a central governing body or funder seeks information from the public and makes executive decisions (Liu et al. 2021). This type of governance is also known as 'consultative' and 'functional levels of participation' (Conrad and Hilchey 2011). The benefits of this approach are standardised protocols and data formats that support interoperability – users and machines know what to expect. Drawbacks include lack of flexibility and challenges of adopting and implementing rigorous standards imposed by the governance body (Ceccaroni, Bowser, and Brenton 2017), valuable knowledge from contributors can be lost if its concepts are not captured by a strictly defined data model.

Bottom-up governance structure often results from a community response to a crisis, with the intention to initiate government action (Conrad and Daoust 2008). This type of governance is also referred to as transformative, community-based, grassroots, or advocacy (Conrad and Hilchey 2011; Wolff and Muñoz 2021). In a ***bottom-up approach***, standards are loosely defined, and all members can participate equally in decision-making. There are views that this type of governance is more favourable and leads to more sustainable use of resources (Bradshaw 2003). The main benefits are flexibility and natural shaping of standards from diverse community contributions. Flexibility can also be a disadvantage, since communal harmonisation and decision-making are slow, and a non-standardised approach affects interoperability and credibility (Bradshaw 2003) when it is impossible to know what to expect from data. Additionally, funding and platform stability can be challenging to maintain (Bradshaw 2003). While on the opposite sides of the spectrum, data collected using bottom-up approach is complementary to the data collected following a top-down participation governance (Elwood, Goodchild, and Sui 2012) and policymakers should find a balance between two approaches (Marchezini et al. 2017).

The Global Biodiversity Information Facility (GBIF)[3] is an international network that promotes and facilitates free and open access to biodiversity data from across the globe. GBIF was established in 2001 through a Memorandum of Understanding between participating governments, and is now funded by agencies from national governments with voting rights. GBIF accepts data

from diverse sources, including citizen science initiatives such as ***iNaturalist*** and ***eBird*** (also identified in our review).

GBIF facilitates searching for species occurrence data, taxonomic information, and biodiversity datasets. The platform contains over 2.5 billion species occurrence records and over 90 thousand datasets. Data is available for download as a zip file in two formats: tab-delimited CSV (only data that has gone through interpretation and quality control), and Darwin Core Archive (DwC-A) (the original data as shared by the publisher(s) and the interpreted quality-controlled data). Each data download has a Digital Object Identifier (DOI) that, in accordance with the licence, must be cited when using the data; this increases transparency and reproducibility by recording the provenance of the data.

While Darwin Core is the required format for GBIF published data, there is consensus that Darwin Core alone is not sufficient to support a variety of richer and more complex types of biodiversity data. GBIF provides Registered Extensions[4] and actively supports the initiative to evolve their biodiversity data model.[5]

The success of GBIF in becoming the largest open biodiversity data provider lies not only in developing a stable software platform but also in providing standardised but evolving data and metadata standards, best practice documents, and technical tools. GBIF ***Darwin Core Archive Assistant***, ***Validator Tool*** and ***Integrated Publishing Toolkit*** facilitate the structuring of data using the DwC-A format, validation of datasets before uploading to GBIF, and publishing of datasets through the GBIF network. These resources make the platform more accessible to a wide range of stakeholders, and ensure data openness, correctness, and interoperability.

GBIF, which includes citizen science observations data, is actively working towards observing FAIR principles and only accepts data contributions that align with FAIR. While GBIF is a potential biodiversity data source for the GDDS, some limitations should be noted. The DwC-A data model ensures that data consumers always know how to query data and what format to expect when downloading it, but also results in the loss of valuable data that does not conform to the model's structure, such data needs to be hosted elsewhere, contributing to data fragmentation. Differentiating citizen science data on GBIF is not a straightforward task; data can be filtered by the provider but, e.g. museums can contribute both official and citizen science observations.

OpenStreetMap (OSM)[6] is a collaborative platform and project that aims to create an editable, open-access map of the world from contributions by citizens. A community of volunteers from across the globe use GPS devices, aerial imagery, and local knowledge to map and verify various features, including roads, buildings, parks, rivers, and more. The platform is financed by regular donations, intermittent fundraising appeals, and OpenStreetMap Foundation membership, and is currently hosted with support from University College London and other partners.

An in-depth review aimed at readers with little knowledge of OSM is offered by Mooney and Minghini (2017). Here, we summarise the key features and most prominent services and tools that utilise OSM data.

A vast and ever-evolving range of third-party applications, tools, and services are developed using OSM data.[7] Commercial companies use OSM data for mapping services (***Geofabrik***), navigation (***Mapbox***, ***Mapzen***, ***OSMAnd***), live traffic updates and road conditions (***MapQuest***), geospatial analytics (***CampToCamp***). Examples of prevalent free OSM-based services and applications include route planning and navigation for outdoor activities (***Komoot***), cycling infrastructure and cycling route planner (***OpenCycleMap***, ***BBBike***), accessibility information for wheelchair users (***WheelMap***), support for humanitarian and disaster response and mapping of the most vulnerable and disaster-prone areas (***The Humanitarian OpenStreetMap Team HOT***, ***Missing Maps***). Successful applications of OSM data, not only in open source but also in commercial settings, demonstrate its high value as a re-usable resource.

The core function of OSM is to collect, maintain, and distribute an open global geospatial database, rather than to produce cartographic products and maps (Mooney and Minghini 2017). The

OSM conceptual data model of the physical world consists of three basic elements[8]: ***nodes*** that define points in space, ***ways*** that define linear features and area boundaries (polygons and polylines), and ***relations*** that define logical collections between elements. On creation, each element in OSM is assigned a unique identifier that is also linked to its subsequent versions. An element must contain at least one tag that describes its specific properties; this creates structured metadata and adds an essential semantic meaning to each element in the database. There are many resources to guide users in identifying appropriate tags and understanding tag usage (e.g. ***TagInfo***[9]).

There are many ways in which OSM data can be accessed[10]: download of a complete copy of OSM database (updated weekly) or a full OSM editing history[11], regional datasets[12], unfiltered raw data[8], data in GeoJSON format.[13] OSM offers a RESTful Editing API supporting developers and applications in creating, reading, updating, and deleting OSM data programmatically. Such services facilitate easy access to and interoperability of OSM data, a crucial aspect for seamless integration into the GDDS.

The OSM initiative closely aligns with FAIR by observing good practices of open data. Successful applications of OSM data exist covering all the GDDS themes. Some examples include support for global climate resilience[14], studies on urban heat islands (Dimitrov, Popov, and Iliev 2021), classification of local climate zones (Fonte et al. 2019), OSM CircularEconomy project (OSM 2022), environmental assessment studies (Kloog, Kaufman, and Hoogh 2018), research on habitat fragmentation and disturbances (Bista et al. 2021; Snell et al. 2020), crowdsourcing mapathons for detecting deforestation (Bratic and Brovelli 2022), and urban forest mapping (PlanIT 2023). OSM presents a valuable resource for inclusion in the GDDS, though an additional layer of applications and semantic resources will be required to facilitate data discovery and data integration with other sources.

Sensor.Comunity[15], formerly Luftdaten.info, is an open-source, community-driven project aimed at building and deploying low-cost air quality sensors and providing real-time high-resolution air quality data at the local level. Luftdaten.info was established by the Open Knowledge Lab (OK Lab) in Stuttgart in 2015 (re-branded as Sensor.Community in 2019) as a German air quality project, and quickly grew into a global citizen science community (although currently most sensors are concentrated in Europe). The project is supported by volunteers and voluntary donations.

Sensor.Community's goal is to raise awareness about air pollution and its potential health and environmental impacts, enable citizens to actively participate in monitoring and improving air quality in their communities, and to create a comprehensive dataset that can be used for research, advocacy, and policymaking related to air quality improvement. Some example applications of Sensor.Community data include a ***Samen voor Zuivere Lucht***[16] platform that combines Sensor.Community data with official data sources, ***Samen Meten***[17] portal that harvests Dutch data from Sensor.Community database, and ***HackAir***[18] platform that uses Sensor.Community data to generate information on air quality, thermal comfort, and the probability of forest fires in Europe.

The sensor kits can be assembled to measure environmental factors (temperature, pressure, humidity), particulate matter pollutants (PM10 and PM2.5), and noise, and once configured, can be registered with the platform. The aggregated results are displayed on a live map from nearly 13,000 active sensors in 78 countries with over 23 billion data points.

Historic data from 2015 onwards can be downloaded from the Sensor.Community Archive[19] automatically by writing custom scripts. Aggregated daily readings for each sensor are served as CSV files with file names indicating the date, type of sensor, and sensor ID. Sensor kits can contain multiple sensors (environmental, pollutants, and/or noise), each of these sensors will generate a separate CSV file in the historic database. Location information (latitude and longitude) can be used to identify sensors that belong to the same sensor kit. No standardised metadata is currently supplied to describe the sensor readings in the database.

Sensor.Community data is currently free and open, with clearly-documented licence conditions. It fulfils a number of the FAIR principles, and has the potential to complement official

environmental data sources at a local scale in the GDDS context of climate change and pollution. To accomplish successful inclusion within the GDDS, Sensor.Community data (like most sources of air quality information) could be semantically enriched using controlled vocabularies such as DEFRA Air Pollution Glossary, Eionet Data Dictionary, or other to ensure seamless integration with other sources. Additional APIs or service layers could facilitate data search (by sensor ID, date/time location, etc.) and aggregation of measurements from the same sensor kits (e.g. for calibration or data quality estimation).

4.1. Discussion: FAIRness in citizen science data

As discussed earlier in this section, there are two main models of governance of participation – top-down and bottom-up approaches – that define how data FAIRness can be achieved in citizen science. GBIF follows a top-down approach by specifying rigorous standards for data contribution (DwC-A and EML). It observes FAIR by setting metadata and data requirements and assigning DOIs (**F**), offering an API and machine-readable interface (REST + JSON) (**A**), using Ecological Metadata Language (EML) and DwC-A (**I**), requiring creative common data licences, and recording data provenance (**R**).

OSM and Sensor.Community are examples of bottom-up approaches where data structure and documentation emerged from the community contributions. OSM free-text tagging has evolved into a database of community-accepted, commonly used tags. The use of persistent identifiers facilitates the recording of the full history of changes to the nodes (**F**), and various applications provide data search and download capabilities (**A**), consistent data formats support interoperability (**I**), DbCL v1. 0 licence ensures the traceability of data (re-)use (**R**). The structure of Sensor.Community data is defined by the specific sensors used to collect data, but as new sensor kits become available, new data fields will emerge. **I**nteroperability and **R**euse are facilitated using a simple data format (CSV) and by offering data under DbCL v1. 0 licence. Further alignment with FAIR can be achieved by tagging with semantic resources (**F, I**) and developing a search and download API (**A**).

5. Tools that support citizen science

To achieve data FAIRness, projects must follow good practice from the project planning stage and produce (or adopt) a suitable Data Management Plan. However, many citizen science projects may struggle with finding and selecting a compatible set of tools, standards, and protocols to support them with all stages of the project lifecycle. Adding to this challenge, free and open-source tools typically carry several, but not all, functions to deliver a project end-to-end. For instance, the primary role of Zooniverse (discussed in Section 5.2) is project hosting, with additional facilities for basic project search and data storage (for active projects), but no support for data publishing. In this section, we discuss in detail some of the more prominent tools, resources and standards identified in the review to explore the functions such tools can offer; a full summary is in Appendix 1 and 2.

5.1. Planning and conception of data governance

The first step in achieving data FAIRness is a strong Data Management Plan that (among other things) considers participation consent, (meta)data formats, (meta)data standards and vocabularies, data structuring, data documentation, data licensing, data hosting, and data sharing. Most citizen science platforms, network websites, and online tools provide free supporting materials, guides, and/or training courses for citizen science project managers, educators, researchers, citizens, and other stakeholders.

EU-Citizen.Science, for instance, offers training courses including '*FAIR Data in Citizen Science Projects*' and '*Doing Citizen Science as Open Science*' (other resources summarised in Appendix 1).

Advanced search and filtering of such resources is not yet supported, so citizen science stakeholders either need to know what they are looking for or to manually inspect the resources that appear relevant.

5.2. Project hosting

Citizen science projects can either be hosted on a dedicated third-party platform or can develop their own infrastructure for participation and data collection. The latter can be resource-intensive, depending on the project ambition and the complexity of the platform required. Our review identified 12 project hosting platforms which are summarised in Appendix 2. Here, we discuss ***Anecdata***, ***CitSci*** and ***Zooniverse,*** since at present these platforms support the largest number of citizen science projects.

Anecdata[20] is a free community science platform founded in 2014 by the Community Lab at the MDI Biological Laboratory in Bar Harbor, Maine. It is well suited for more complex biodiversity protocols such as recording absence data, water quality monitoring, litter recording and clean up, and collection of non-biodiversity image observations.

Anecdata allows project owners to create projects, define data sheets with multi-dimensional data, select participation mode, share data publicly or keep it private to the project. The platform offers a free mobile app (iOS and Android) to collect observations from the field with support for geoprivacy. Either Creative Commons Attribution 4.0 International License or Open Data Commons Attribution License (ODC-By) v1.0 can be selected for the data collected via Anecdata platform. The platform contains over 300 active projects, 15,500 users, 111,000 observations, and 74,000 photos and images.

Anecdata facilitates access to public observations in a tabular format or displayed on a map. Observation data can be filtered by project name, date range, user who submitted the observations, and location. There is no option to filter observations by tags, which limits the ability to obtain all observations for the required domain or topic.

CitSci[21] offers free tools for the entire citizen science project process, from project creation, management of participants, building custom data sheets to collecting data, analysing data, sharing data, and gathering community feedback. Observations can be added via a web form or CitSci mobile app (Android and iOS). Only project members can contribute data; memberships can be open (any registered user can join without owner approval) or closed (requests to join require owner approval). Project owners can permit project members or the public to view data in tabular format or on a map and download data in Excel or CSV format. Most projects choose to restrict data downloads to members only, or entirely disallow downloads. The platform hosts 1,133 projects and has 147,504 observations; most projects are located in the US. While projects created and hosted on CitSci can be automatically published to SciStarter for discovery, it is not possible to search or access data in a straightforward manner.

Zooniverse[22], former Galaxy Zoo, is a free platform designed for projects that need volunteer support in classifying or annotating images, transcribing historical documents, identifying patterns in data, and other classification tasks. Projects can create tutorials, define workflows (sequences of tasks), set questions and drawing tasks, and more. Zooniverse lists 97 active, 243 paused, and 110 finished projects, but the project search is limited to the domain and project name. Completed projects can publish aggregated results and reports; however, data downloads are not supported.

Anecdata and CitSci may appear as potential citizen science data sources for the GDDS; however, in practice, their main function is project hosting with limited data discovery for reuse.

5.3. Data collection

The requirements for data collection will vary based on the nature of the project. Observation tasks will typically require tools that support custom datasheets, multimedia upload, mobile apps or

mobile-friendly web interfaces, and secure data transactions. For sensor data like air or water quality measurements, the management and retrieval of observations and metadata and the implementation of secure Internet of Things (IoT) protocols become essential. Enriching the data will involve annotation, classification, or workflows.

Our review identified 23 data collection platforms which are summarised in Appendix 2. Some of these platforms, e.g. iNaturalist, eBird, GLOBE Observer, cannot be customised and can only be used to contribute data to their corresponding initiatives. While Natusfera[23] presents an example of interface customisation of iNaturalist, the data is contributed to iNaturalist platform. Here, we discuss ***ODK*** and ***ArcGIS tools*** as examples of customisable free and commercial data collection platforms widely applied in citizen science.

Open Data Kit (ODK)[24] is designed for building custom data collection forms on mobile devices to support efficient and reliable data collection, especially in offline or low-connectivity environments. ODK is widely used in public health, humanitarian aid, environmental monitoring, and social research (Hartung et al. 2010; Tom-Aba et al. 2015; Campus et al. 2020). The three key components of the ODK platform are ***ODK Collect*** (an Android application for building custom data collection forms and capturing data), ***ODK Build*** (no-code web-based survey designer tool for customised forms), and ***ODK Central*** (the ODK server that acts as a central repository).

ArcGIS[25] is a commercial cloud-based software toolkit for capturing, managing, analysing, and displaying geospatial data which is used by a variety of citizen science projects (e.g. Hawthorne et al. 2015; Spear, Pauly, and Kaiser 2017; Chmielewski et al. 2018). ***ArcGIS Survey123***[26] offers a fully customisable survey product for data collection via a web browser or mobile application. Data collection forms can include lines and polygons, images and audio files, high-accuracy data capture. ***ArcGIS QuickCapture***[27] survey product for field observations allows capture of images and sensor information from devices on moving vehicles. ***ArcGIS Community Science Solution***[28] is specifically designed for collecting location-enabled plant and animal observations from citizen scientists and is primarily used by conservation organisations, natural resource departments, and other government agencies.

There are important tradeoffs to consider between cost and technical capacity: open-source tools such as ODK offer free data collection capabilities, but require technical competency and a private or cloud-based server to run the software code and store data. Commercial ArcGIS solutions provide flexible off-the-shelf data collection capabilities; however, these can be costly for small-scale citizen science projects.

5.4. Semantic resources

Semantic resources are essential for data FAIRness: (meta)data standards, controlled vocabularies or other structured data descriptions (e.g. data tagging) facilitate data discovery, interoperability, (re)use, and integration (especially across domains). Such resources could be integrated within citizen science project hosting platforms to offer pre-populated lists of terms for creating datasheets (with an option for customisation) rather than every project defining its own vocabularies. For instance, CitSci does not endorse any semantic resources (datasheet templates are under development), which results in an unpredictable data structure, ultimately impacting interoperability.

Our review identified 8 semantic resources relevant to the biodiversity, environment, bioinformatics, oceanography, and agriculture domains (summarised in Appendix 2). Here, we discuss Darwin Core, EnvO, and NERC Vocabulary Server to exemplify functions that controlled semantic resources can offer to users.

Darwin Core[29] encompasses two functions: an evolving semantic resource, and a structural data standard for publishing, integrating, and sharing biodiversity information. Darwin Core contains a glossary of terms and '*is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information*'. Since Darwin Core carries two functions, we will continue the discussion in Section 5.7.

The Natural Environment Research Council (NERC) Vocabulary Server (NVS)[30] is a collection of standardised and hierarchically structured controlled vocabularies, primarily covering oceanographic and related domains with example applications in citizen science (Busch et al. 2016). The platform comprises vocabularies and thesauri stored as Linked Data in human – and machine-readable formats. NVS supports basic searches based on simple text matching, advanced searches for terms in specified vocabularies or across vocabulary collections, and interrogation of mappings between different vocabularies. An ***Interactive Query UI***[31] provides a simple interface to query NVS triplestore (the RDF database of all NVS vocabularies) using SPARQL queries. It also allows automatic encoding of SPARQL queries into a single line string and decoding back into SPARQL queries format.

The Environment Ontology (EnvO)[32] is a FAIR-compliant community ontology that offers concise, controlled description of environments from microscopic to intergalactic scales. EnvO was established in 2013 as a simple ontology and grew with the support of the ESIP Federation, UN Environment, IOC-UNESCO, and individual contributions. It contains over 7,000 classes and allows requests for new terms and synonyms, enhancements, or reporting defects via GitHub issue tracker.[33] Subsets of terms linked to the EnvO Internationalised Resource Identifiers (IRIs) for traceability can be generated to tailor particular needs. Subsets can be hosted for projects or communities on EnvO GitHub on request. The ontology can be downloaded from OntoBee web server[34], EBI Ontology Lookup Service repository[35], or EnvO GitHub repository.[36]

Semantic resources are increasingly used by the scientific community (Leadbetter 2015; Magagna et al. 2021) but as yet are rarely considered by citizen science initiatives. One factor contributing to this is the limited awareness of available resources and their role in data interoperability (and the importance of interoperability itself). Certain semantic resources might be overly complex for citizen science initiatives, but relevant terms can be extracted into custom vocabularies or ontologies and referenced back to the original sources, e.g. using Semantic Treehouse vocabulary hub (Van den Berg 2023). Tools like OntoPortal[37] can be used to support citizen science communities in building ontology repositories, annotating free text with the vocabulary terms, identifying the associations between terms, and offering recommendations on semantic resources.

5.5. Data publishing and preservation

To ensure long-term value outside the project that collected data (and to ensure FAIRness), data needs to be hosted in an accessible manner. Those projects which do publish their data may use their own infrastructure, which makes data difficult to discover. Schade and Tsinaraki (2016) revealed that the majority of surveyed projects host their data on a remote server (38%) or a local machine (16%) managed by a project member. However, it remains unclear whether this data is catalogued and is discoverable elsewhere. Other projects collect data suited for contribution to larger initiatives that already provide open data capabilities, e.g. iNaturalist, eBird, GLOBE, and Sensor.Community. Ideally, data (or reference to data) that does not fit domain-specific platforms should be published on a suitable platform so that it can be easily discovered, acquired, and (re)used. Our review identified two data repositories used by citizen science projects: ***Zenodo*** and ***Mendeley Data***.

Zenodo[38] is a multidisciplinary open repository designed for research communities to deposit research datasets, software, reports, papers, and other digital research artifacts. The platform was launched in 2013 and is owned by the European Organization for Nuclear Research (CERN). Registered users can deposit research artifacts under closed, open, or embargoed access and at any stage of the research lifecycle, provided that they hold appropriate rights for the materials. Zenodo offers a RESTful API to support deposit of research outputs, records search, and files upload and download. All uploads are assigned a DOI for traceability.

Zenodo integrates with other research platforms and services, including ***GitHub*** for automatic synchronisation between code repositories and associated research outputs; ***ORCID*** (Open

Researcher and Contributor ID) to connect researchers' ORCID profiles to Zenodo, ensuring attribution and recognition for their deposited research outputs; ***DataCite*** to provide persistent identifiers; ***OpenAIRE*** (Open Access Infrastructure for Research in Europe) to index Zenodo content in the OpenAIRE[39] database, enhancing discoverability and accessibility within the open science community; ***CERN Analysis Preservation (CAP)***[40] infrastructure enabling researchers to preserve and share their analysis workflows, code, and associated data in a FAIR manner.

Mendeley Data[41] is a free multidisciplinary open repository designed for long-term data storage. The platform is a product of Elsevier and was launched in 2016. Mendeley Data fully supports FAIR principles (Elsevier 2020) however it is an institutional data repository and is only available to registered research institutions. All datasets (including the underlying assets and versions) include deep-indexing of both metadata and files; metadata is indexed in common search indexes, such as Google Dataset Search, DataCite Search, OpenAIRE with OAI-PMH, and Share from Open Science Framework. Artifacts can be deposited under closed, open, or embargoed access. Mendeley Data offers a Digital Commons Data API for managing and searching of research artefacts. The platform supports standard metadata schema such as Dublin Core and schema.org, controlled vocabularies for standard fields and custom metadata fields which can be configured to use values from existing taxonomies for interoperability, discoverability and reuse.

Zenodo and Mendeley Data support advanced search by constructing complex text-based queries, though discovering new relevant geospatially tagged resources and datasets can be extremely challenging. Both platforms support dataset updates and DOI versioning, but it is impractical to generate an excessive number of versions. This is a potential limitation for hosting data from long-term or ongoing citizen science projects that generate continuously evolving datasets, rather than static data snapshots or regular 'releases'.

5.6. Standards for structuring and accessing data

International data standards ensure consistency and interoperability among the data collected by different individuals or groups participating in citizen science initiatives (Schade et al. 2017; Bowser et al. 2020). Their use enhances the credibility and scientific value of citizen science efforts (Spasiano et al. 2021), making the data more reliable for researchers, policymakers, and the broader community. Standards also facilitate collaboration and knowledge sharing, ultimately contributing to the success and impact of citizen science projects.

Our review identified 11 standards detailed in Appendix 2. Here, we discuss ***Darwin Core*** (introduced earlier), the ***Open Geospatial Consortium (OGC) Observations and Measurements (O&M)*** (fundamental to the STA standard identified in our review), and ***OGC Sensor Things API (STA)*** as examples of generic data structure and data sharing standards that are relevant to citizen science, and ***STAplus*** and ***PPSR Core*** as standards that are specifically designed for citizen science initiatives.

Darwin Core was developed by the Biodiversity Information Standards (TDWG) community and ratified as a standard in October 2009 (Wieczorek et al. 2012). It is based on the Dublin Core, Species Analyst[42], and the Access to Biological Collections Data (ABCD) standards, designed to be minimal (only to include essential terms) and, unlike ABCD, it is flat, i.e. with no relational structure. The standard is maintained in RDF but is available in the HTML, RDF/Turtle, RDF/XML, and JSON-LD formats. As discussed in Section 4, Darwin Core is the preferred standard for publishing data to GBIF.

The ***OGC O&M / ISO 19156:2011*** is an international standard that defines a conceptual framework and encoding for describing observations and measurements. It provides a standardised way to model and exchange information about various types of observations from sensors, instruments, algorithms, or process chains. In the context of the O&M, citizen scientists could be seen as instruments or sensors themselves that collect observations about a phenomenon.

The O&M data model is fundamental as the core of OGC Sensor Web Enablement (SWE) standards such as SensorThings API, WaterML 2.0, and Sensor Observation Service (SOS). It defines a core set of properties for observing a phenomenon (Figure 2): *Feature* (an abstraction of a real-world phenomenon), *Observation* (the act of measuring or obtaining information about a phenomenon), *Feature of Interest* (the entity for which the observation is being made), *Observed Property* (a characteristic, attribute, or property of the phenomenon being observed, e.g. particulate matter in measuring air quality), *Procedure* (the method or process used to make an observation, e.g. instruments, sensors, human observers), and *Result* (the data obtained from an observation, e.g. a single value, a time series, an image).

The Environmental Monitoring Facilities (EMF) data model[43] is an example application of the O&M standard. EMF describes each facility as a spatial object in the context of INSPIRE[44] and links observations and measurements of environmental parameters to the facility, where citizen science is included as one of the stakeholder initiatives for sharing public data.

OGC SensorThings API 1.1 (STA)[45] provides an open and unified way to interconnect heterogeneous Internet of Things (IoT) devices, data, and applications over the Web. The first version 1.0 was published in 2016 (latest version 1.1 in 2021) and developed by the OGC Sensor Web for IoT Standards Working Group (SW-IoT SWG). The standard is designed for organisations that need web-based platforms to manage, store, share, and analyse IoT-based sensor observation data across domains.

The key entities specific to STA are (Internet of) *Thing*, defined as '*an object of the physical world (e.g. device) or the information world (e.g. system) that is capable of being identified and integrated into communication networks*' (ITU 2012), as well as associated *Location* and *Datastream* (a collection of observations from a single sensor) (Figure 3). Entities, such as *FeatureOfInterest*, *ObservedProperty*, and *Observation* are based on the OGC O&M model.

STA is relevant for the GDDS particularly because of its increasing use in IoT platforms for environmental monitoring and smart cities, including the FROST Server open source implementation of STA[46], an STA-based INSPIRE download service[47], and the adoption of STA by the French Geological Survey[48] for the national groundwater monitoring system and water quality database.

OGC SensorThings API Extension: STAplus 1.0[49] is an approved international standard and an extension of the STA data model based on requirements from the citizen science community. FAIR (in particular, Interoperable and Reusable) principles are reinforced by adding entities of ownership, licence, and project information for sharing observations. The extension also enables users to express explicit relations between observations and to create group(s) of observations that belong together.

The STAplus data model describes five entities in addition to the STA (Figure 4): *Party* (links a user to a Datastream or Group), *License* (specifies reuse conditions), *Project* (allows for organising a

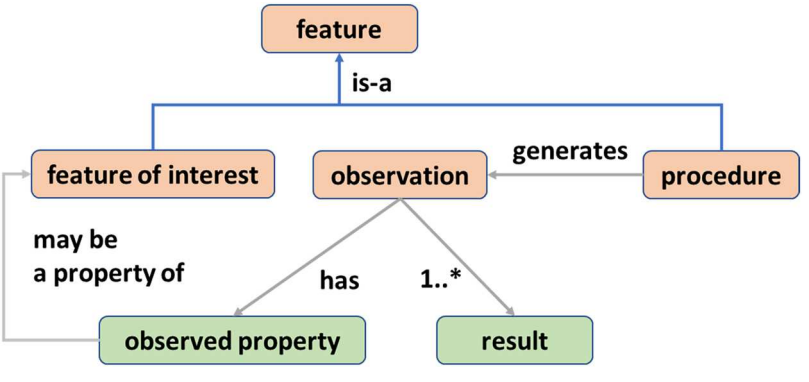

Figure 2. Basic structure of the OGC Observations and Measurement Model (adopted from Usländer, Coene, and Marchetti 2012).

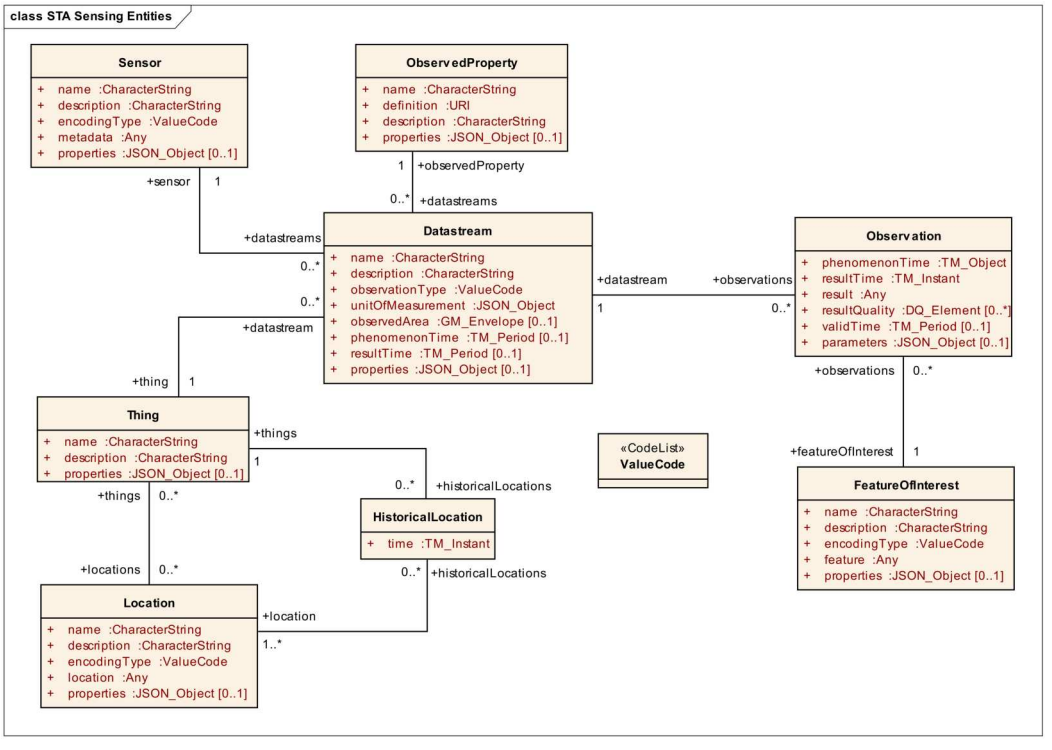

Figure 3. Sensing entities of SensorThings API.

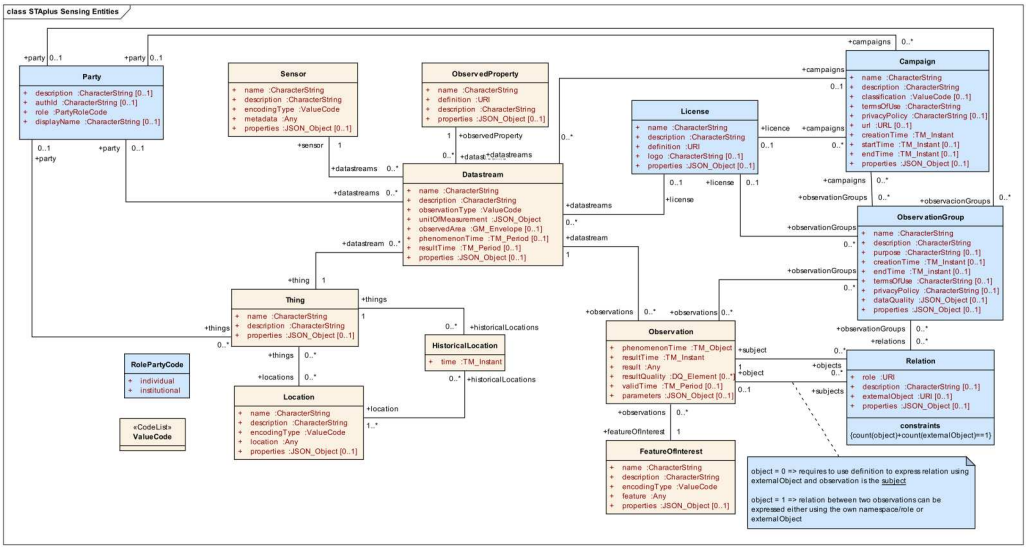

Figure 4. Sensing entities of STAplus.

campaign or project), *Group* (allows to package individual Observations as a bag or set), and *Relation* (supports relationships between Observations).

The '*OGC Best Practice for using SensorThings API with Citizen Science*' document[50] offers practical examples of applying the STAplus extension in the citizen science domain.

PPSR Core[51] is an open data and metadata standard that defines a common framework for describing citizen science projects. The PPSR Core initiative started in 2013, supported by the DataONE PPSR Working Group and SciStarter. It is now maintained by the Citizen Science Association's Data & Metadata Working Group with support from volunteers. The standard is still under development but is designed to enable the sharing of basic common information across databases that catalogue citizen science projects. It facilitates consistent project discovery between all major project discovery platforms including SciStarter, CitSci, Atlas of Living Australia BioCollect, and CitizenScience.gov.

The PPSR Core standard comprises four models: *Common Data Model (CDM)* for aggregating citizen science projects into programs or campaigns within a common organising framework, *Project Metadata Model (PMM)* for describing the purpose, responsible parties, participation and engagement, and other contextual information for citizen science projects, *Dataset Metadata Model (DMM)* for describing collections of observations (e.g. protocols, temporal range, licence), *Observation Data Model (ODM)* for defining domain 'profiles', i.e. core sets of features that should be collected for a given study. PMM includes some controlled vocabularies, but projects are welcome to adopt other semantic resources, provided that they are clearly referenced.

Foundational data standards, such as O&M and STA, can serve as the basis for tailored extensions to meet the needs of citizen science initiatives. Clear supporting documentation of best practice (e.g. OGC 2022) plays an important role in providing use cases and improving understanding of how standards can be applied in practice. Additionally, for services based on APIs (e.g. STA FROST Server), tools similar to the NVS Interactive Query UI could be developed, to offer a user-friendly interface for constructing complex API queries and encoding these as URLs.

6. Discussion and conclusions

The importance of FAIR is increasingly being acknowledged within the field of citizen science, as demonstrated by the major citizen science initiatives promoting FAIR (EU-Citizen.Science 2021; ECSA 2023), and recent research into citizen science data FAIRification (Coché et al. 2021; Ramírez-Andreotta et al. 2021; Turicchia et al. 2021b; Alvarez et al. 2022). However, citizen science projects that operate independently from larger initiatives may still lack awareness of FAIR principles, struggle to select suitable standards and tools, or fail to recognise the value of sharing their data outside of the project.

Our review identified a number of tools that can facilitate different stages of the citizen science project lifecycle. Commercial solutions like ArcGIS and SPOTTERON offer a full suite of tools and applications to support an end-to-end data lifecycle; however, these may be costly for smaller-scale projects. Free and open-source platforms and tools are generally limited in functionality required for end-to-end data lifecycle management in a FAIR way. Therefore, projects might need to select and combine different tools by purpose from different providers, resulting in more challenges to achieve a seamless flow of FAIR data.

On the face of it, it may seem paradoxical that commercially-licensed software and platforms are discussed in the context of FAIR data. While 'FAIR' does not necessarily equate to 'open' (Jeffery 2021) FAIR data are required to have clear licence information, ideally in machine-readable form, and citizen science data governance involves obtaining and documenting the consent of contributors for their data to be used and re-used in specific contexts. Any technical tools which assist in this governance might be considered as assisting on the path towards FAIR data. This is also an important consideration in the EU GDDS, which of necessity will bring together commercial and private stakeholders with public sector players, requiring 'transparent but controlled accessibility of data and services' (Mons et al. 2017).

While a vast number of domain-specific controlled vocabularies and other semantic resources exist, our review identified only 8 semantic resources used by the citizen science projects. This may indicate that independent projects rarely apply standardised semantic tagging because they

are either unaware of its importance or unsure which resources to choose from a confusing range. This presents a major gap in data discovery and interoperability, especially cross-domain. In addition, as demonstrated by Ramírez-Andreotta et al. (2021), projects may have to create custom semantic resources by combining subsets of controlled vocabularies and introducing new custom terms to fulfil their needs. Tools for selecting and extending semantic resources, similar to the EcoPortal[52] tools which practically implement an OntoPortal for the ecological domain, need to be developed for a wide range of domains to support the citizen science community (de Sherbinin et al. 2021).

The availability of centralised data repositories for citizen science data presents another major challenge. Platforms created during time-limited research projects may not be accessible after project funding terminates. Open repositories such as Zenodo and Mendeley can be used to publish and share citizen science datasets, but search capabilities are limited (e.g. it is not straightforward to filter citizen science data). Another limitation is that such platforms only facilitate publishing of static datasets which might be suited for completed projects; dynamic projects will need to publish periodic 'snapshots' of their data. The quality of data collected by citizen science projects may be in question when the methodology is not transparently documented or robust. If a data repository platform is tailored for citizen science, citizen science data can be improved by AI technologies and validated by expert knowledge, as practised in the iNaturalist and eBird platforms.

The successful development of the EU GDDS will depend on the availability of FAIR data sources, including FAIR citizen science data. Large longstanding initiatives such as GBIF already offer FAIR data that can be easily integrated within the GDDS. Other large community platforms such as OpenStreetMap and Sensor.Community will require additional layers of tools and semantic resources to enable integration. Smaller-scale projects with limited resources may miss opportunities to offer their data for re-use in the absence of free end-to-end solutions to support production and sharing of FAIR data. This presents a major challenge for the GDDS to deliver the ambition of establishing a Single EU Market for data and integrating citizen contributions as defined by the Open Science Policy of the EC.

FAIRification of citizen science data has significant importance beyond policy making and decision support. Adherence to FAIR principles can improve knowledge mobilisation, strengthening capacity to conduct research using citizen science data. Production of FAIR data can also help empower communities by making their data more visible to and accessible by authorities. It can additionally increase community engagement, e.g. a case study on flood monitoring (Wolff 2021) showed that community members highly value ability to access and share their data.

There is the potential for citizen science data to be integrated in environmental Research Infrastructures[53] or e-infrastructures that can serve as intermediaries connecting to the GDDS and supporting data sharing. In exchange for citizen science data, such infrastructures should increase technical and semantic services to facilitate citizen science projects in meeting high Technological Readiness Levels (Mankins 1995) in operational environments. Future citizen science project calls should include a strategic plan on how services developed during the project period will be sustained by connecting them with specific environmental Research Infrastructures such as LifeWatch ERIC, eLTER, or others from the environmental cluster of RIs.

Notes

1. https://eu-citizen.science/.
2. https://scistarter.org
3. https://www.gbif.org
4. https://rs.gbif.org/extensions
5. https://www.gbif-uat.org/composition/HjlTr705BctcnaZkcjRJq/gbif-new-data-model
6. https://www.openstreetmap.org
7. https://wiki.openstreetmap.org/wiki/List_of_OSM-based_services
8. https://wiki.openstreetmap.org/wiki/Elements

9. https://taginfo.openstreetmap.org/
10. https://wiki.openstreetmap.org/wiki/Downloading_data
11. https://planet.openstreetmap.org/
12. https://download.geofabrik.de/
13. https://wiki.openstreetmap.org/wiki/Overpass_turbo/GeoJSON
14. https://www.hotosm.org/updates/how-hot-is-supporting-climate-resilience-around-the-world/
15. https://sensor.community/en/
16. https://samenvoorzuiverelucht.eu/dataportaal/
17. https://samenmeten.rivm.nl/dataportaal/
18. https://platform.hackair.eu/
19. https://archive.sensor.community/
20. https://anecdata.org
21. https://www.citsci.org/
22. https://www.zooniverse.org
23. https://cos4cloud-eosc.eu/citizen-science-innovation/cos4cloud-citizen-observatories/natusfera/
24. https://getodk.org/
25. https://www.arcgis.com/index.html
26. https://survey123.arcgis.com/
27. https://www.esri.com/en-us/arcgis/products/arcgis-quickcapture/overview
28. https://doc.arcgis.com/en/arcgis-solutions/latest/reference/introduction-to-community-science.htm
29. https://dwc.tdwg.org/
30. https://vocab.nerc.ac.uk
31. http://vocab.nerc.ac.uk/sparql
32. https://sites.google.com/site/environmentontology/home
33. https://github.com/EnvironmentOntology/envo/issues/Ontology/envo
34. https://ontobee.org/ontology/ENVOVO
35. https://www.ebi.ac.uk/ols4/ontologies/envoService
36. https://github.com/EnvironmentOntology/envo/
37. https://ontoportal.org/
38. https://zenodo.org/
39. https://explore.openaire.eu
40. https://analysispreservation.cern.ch
41. https://data.mendeley.com/
42. http://xml.coverpages.org/speciesAnalyst.html
43. https://inspire.ec.europa.eu/theme/ef
44. http://www.datacove.eu/inspire/o-m-in-inspire-sdi
45. https://www.ogc.org/standard/sensorthings/
46. https://www.iosb.fraunhofer.de/en/projects-and-products/frost-server.html
47. https://inspire.ec.europa.eu/good-practice/ogc-sensorthings-api-inspire-download-service
48. https://www.brgm.fr/en
49. https://docs.ogc.org/DRAFTS/22-022.html
50. https://docs.ogc.org/bp/21-068.pdf
51. https://core.citizenscience.org
52. https://ecoportal.lifewatch.eu/
53. https://roadmap2021.esfri.eu/media/1259/rm21-part-2-env.pdf

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work has been co-funded by the European Union, Switzerland and the United Kingdom under the AD4GD project.

Data availability statement

Data sharing is not applicable to this article, as no new data were created or analyzed in this study.

ORCID

Victoria Lush http://orcid.org/0000-0003-3248-9608
L. Bastin http://orcid.org/0000-0003-1321-0800
K. Otsu http://orcid.org/0000-0001-7835-0949
J. Masó http://orcid.org/0000-0002-2983-4629

References

Aceves-Bueno, Eréndira, Adeyemi S. Adeleye, Marina Feraud, Yuxiong Huang, Mengya Tao, Yi Yang, and Sarah E. Anderson. 2017. "The Accuracy of Citizen Science Data: A Quantitative Review." *The Bulletin of the Ecological Society of America* 98 (4): 278–290. https://doi.org/10.1002/bes2.1336.

Althaus, F., N. Hill, R. Ferrari, L. Edwards, R. Przeslawski, C. H. L. Schönberg, R. Stuart-Smith, et al. 2015. "A Standardised Vocabulary for Identifying Benthic Biota and Substrata from Underwater Imagery: The CATAMI Classification Scheme." *PLoS One* 10 (10): e0141039. https://doi.org/10.1371/journal.pone.0141039.

Alvarez, Reynaldo, César González-Mora, José Zubcoff, Irene Garrigós, Jose-Norberto Mazón, and Hector Raúl González Diez. 2022. "FAIRification of Citizen Science Data Through Metadata-Driven Web API Development." *Proceedings of the International Conference on Web Engineering* 13362: 162–176.

Bäckstrand, Karin. 2003. "Civic Science for Sustainability: Reframing the Role of Experts, Policy-Makers and Citizens in Environmental Governance." *Global Environmental Politics* 3 (4): 24–41. https://doi.org/10.1162/152638003322757916.

Balázs, Bálint, Peter Mooney, Eva Nováková, Lucy Bastin, and Jamal Jokar Arsanjani. 2021. "Data Quality in Citizen Science." *The Science of Citizen Science* 139.

Bista, Damber, Greg S. Baxter, Nicholas J. Hudson, Sonam Tashi Lama, and Peter John Murray. 2021. "Effect of Disturbances and Habitat Fragmentation on an Arboreal Habitat Specialist Mammal Using GPS Telemetry: A Case of the Red Panda." *Landscape Ecology* 37: 1–15.

Bonney, Rick, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk. 2009. "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy." *BioScience* 59 (11): 977–984. https://doi.org/10.1525/bio.2009.59.11.9.

Bowser, Anne, Caren Cooper, Alex De Sherbinin, Andrea Wiggins, Peter Brenton, Tyng-Ruey Chuang, Elaine Faustman, Mordechai Haklay, and Metis Meloche. 2020. "Still in Need of Norms: The State of the Data in Citizen Science." *Citizen Science: Theory and Practice* 5 (1).

Bradshaw, Ben. 2003. "Questioning the Credibility and Capacity of Community-Based Resource Management." *Canadian Geographies / Géographies Canadiennes* 47 (2): 137–150. https://doi.org/10.1111/1541-0064.t01-1-00001.

Bratic, G., and M. A. Brovelli. 2022. "Crowdsourcing for Deforestation Detection in the Amazon." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B4-2022: 231–238. https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-231-2022.

Busch, Julia A., Raul Bardaji, Luigi Ceccaroni, Anna Friedrichs, Jaume Piera, Carine Simon, Peter Thijsse, Marcel Wernand, Hendrik J. Van der Woerd, and Oliver Zielinski. 2016. "Citizen Bio-Optical Observations from Coast- and Ocean and Their Compatibility with Ocean Colour Satellite Measurements." *Remote Sensing* 8 (11): 879. https://doi.org/10.3390/rs8110879.

Campus, Sergio Francesco, Roberto Scotti, Irene Piredda, Ilenia Murgia, Antonio Ganga, and Filippo Giadrossich. 2020. "The Open Data Kit Suite, a Mobile Data Collection Technology as an Opportunity for Forest Mensuration Practices." *Ann. Silvic. Res* 44: 86–94.

Ceccaroni, Luigi, Anne Bowser, and Peter Brenton. 2017. "Civic Education and Citizen Science: Definitions, Categories, Knowledge Representation." In *Analyzing the Role of Citizen Science in Modern Research*, 1–23. IGI Global.

Chmielewski, Szymon, Marta Samulowska, Michał Lupa, Danbi Lee, and Bogdan Zagajewski. 2018. "Citizen Science and WebGIS for Outdoor Advertisement Visual Pollution Assessment." *Computers, Environment and Urban Systems* 67: 97–109. https://doi.org/10.1016/j.compenvurbsys.2017.09.001.

Coché, Lorraine, Elie Arnaud, Laurent Bouveret, Romain David, Eric Foulquier, Nadège Gandilhon, Etienne Jeannesson, et al. 2021. "Kakila Database: Towards a FAIR Community Approved Database of Cetacean Presence in the Waters of the Guadeloupe Archipelago, Based on Citizen Science." *Biodiversity Data Journal* 9.

Conrad, Catherine T., and Tyson Daoust. 2008. "Community-Based Monitoring Frameworks: Increasing the Effectiveness of Environmental Stewardship." *Environmental Management* 41: 358–366. https://doi.org/10.1007/s00267-007-9042-x.

Conrad, Cathy C., and Krista G. Hilchey. 2011. "A Review of Citizen Science and Community-Based Environmental Monitoring: Issues and Opportunities." *Environmental Monitoring and Assessment* 176: 273–291. https://doi.org/10.1007/s10661-010-1582-5.

Council of the EU. 2021. "Conclusions on the future governance of the European Research Area (ERA)." Accessed August 30, 2023. https://data.consilium.europa.eu/doc/document/ST-14126-2021-INIT/en/pdf.
DEIMS. 2023. "Welcome to DEIMS-SDR." Accessed August 30, 2023. https://deims.org/.
de Sherbinin, Alex, Anne Bowser, Tyng-Ruey Chuang, Caren Cooper, Finn Danielsen, Rorie Edmunds, Peter Elias, et al. 2021. "The Critical Importance of Citizen Science Data." *Frontiers in Climate* 3: 20.
Dimitrov, Stelian, Anton Popov, and Martin Iliev. 2021. "An Application of the LCZ Approach in Surface Urban Heat Island Mapping in Sofia, Bulgaria." *Atmosphere* 12 (11): 1370. https://doi.org/10.3390/atmos12111370.
EC. 2019. "Open Science." Accessed August 30, 2023. https://research-and-innovation.ec.europa.eu/system/files/2019-12/ec_rtd_factsheet-open-science_2019.pdf.
EC. 2020. "European Green Deal Call: €1 billion investment to boost the green and digital transition." Accessed August 30, 2023. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1669.
EC. 2022. "Staff working document on data spaces." Accessed August 30, 2023. https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces.
EC. 2023. "A European Green Deal Striving to be the first climate-neutral continent." Accessed August 30, 2023. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.
ECSA. 2015. "Ten Principles of Citizen Science". Berlin. Accessed August 30, 2023. https://doi.org/10.17605/OSF.IO/XPR2N
ECSA. 2023. "ECS | European Citizen Science". Accessed August 30, 2023. https://www.ecsa.ngo/cases/eu-citizen-science/.
Eitzel, Melissa, Jessica Cappadonna, Chris Santos-Lang, Ruth Duerr, Sarah Elizabeth West, Arika Virapongse, Christopher Kyba, et al. 2017. "Citizen Science Terminology Matters: Exploring key Terms." *Citizen Science: Theory and Practice*, 1–20.
Elsevier. 2020. "Making your data FAIR with Mendeley Data". Accessed August 30, 2023. https://assets.ctfassets.net/zlnfaxb2lcqx/1i7e6L9fWtPoEMynm0brai/a299816711e4f07ee72d816d1a287e6d/ACAD_RES_MDD_INFO_Fair-data-with-Mendeley-Data_WEB.pdf.
Elwood, Sarah, Michael F. Goodchild, and Daniel Z. Sui. 2012. "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice." *Annals of the Association of American Geographers* 102 (3): 571–590. https://doi.org/10.1080/00045608.2011.595657.
EU-Citizen.Science. 2021. "FAIR data in citizen science projects". Accessed August 30, 2023. https://eu-citizen.science/resource/159.
Farrell, Eimear, Marco Minghini, Alexander Kotsev, Josep Soler Garrido, Brooke Tapsall, Marina Micheli, Monica Posada Sanchez, et al. 2023. "European Data Spaces-Scientific Insights Into Data Sharing and Utilisation at Scale." *Joint Research Centre (Seville Site)*, https://doi.org/10.2760/301609.
Ferri, Michele, Uta Wehn, Linda See, Martina Monego, and Steffen Fritz. 2020. "The Value of Citizen Science for Flood Risk Reduction: Cost–Benefit Analysis of a Citizen Observatory in the Brenta-Bacchiglione Catchment." *Hydrology and Earth System Sciences* 24 (12): 5781–5798. https://doi.org/10.5194/hess-24-5781-2020.
Fonte, Cidalia C., Patrícia Lopes, Linda See, and Benjamin Bechtel. 2019. "Using OpenStreetMap (OSM) to Enhance the Classification of Local Climate Zones in the Framework of WUDAPT." *Urban Climate* 28: 100456. https://doi.org/10.1016/j.uclim.2019.100456.
Fraisl, Dilek, Gerid Hager, Baptiste Bedessem, Margaret Gold, Pen-Yuan Hsing, Finn Danielsen, Colleen B. Hitchcock, et al. 2022. "Citizen Science in Environmental and Ecological Sciences." *Nature Reviews Methods Primers* 2 (1): 64. https://doi.org/10.1038/s43586-022-00144-4.
Fritz, Steffen, Cidália Costa Fonte, and Linda See. 2017. "The Role of Citizen Science in Earth Observation." *Remote Sensing* 9 (4): 357. https://doi.org/10.3390/rs9040357.
GBIF. 2023c. "Diversifying the GBIF Data Model." Accessed August 30, 2023. https://www.gbif-uat.org/composition/HjlTr705BctcnaZkcjRJq/gbif-new-data-model.
GO FAIR. 2016. "FAIR Principles." Accessed August 30, 2023. https://www.go-fair.org/fair-principles/.
Goodchild, Michael F. 2007. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 69: 211–221. https://doi.org/10.1007/s10708-007-9111-y.
Gura, Trisha. 2013. "Citizen Science: Amateur Experts." *Nature* 496 (7444): 259–261. https://doi.org/10.1038/nj7444-259a.
Haklay, Mordechai Muki, Daniel Dörler, Florian Heigl, Marina Manzoni, Susanne Hecker, and Katrin Vohland. 2021. "What is Citizen Science? The Challenges of Definition." *The Science of Citizen Science* 13.
Haklay, Mordechai, Suvodeep Mazumdar, and Jessica Wardlaw. 2018. "Earth Observation Open Science and Innovation." *Earth Observation Open Science and Innovation*, 69–88. https://doi.org/10.1007/978-3-319-65633-5_4.
Hartung, Carl, Adam Lerer, Yaw Anokwa, Clint Tseng, Waylon Brunette, and Gaetano Borriello. 2010. "Open Data Kit: Tools to Build Information Services for Developing Regions." In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, 1–12.
Hawthorne, T. L., V. Elmore, A. Strong, P. Bennett-Martin, J. Finnie, J. Parkman, T. Harris, J. Singh, L. Edwards, and J. Reed. 2015. "Mapping Non-Native Invasive Species and Accessibility in an Urban Forest: A Case Study of

Participatory Mapping and Citizen Science in Atlanta, Georgia." *Applied Geography* 56: 187–198. https://doi.org/10.1016/j.apgeog.2014.10.005.
INSPIRE. 2022. "The development of a European Common Green Deal data space." Accessed August 30, 2023. https://wikis.ec.europa.eu/download/attachments/68190468/DOC3_16MIG_GreenDealDataSpace.pdf?version=1&modificationDate=1669068487597&api=v2.
ITU. 2012. "Overview of the Internet of Things." Accessed August 30, 2023. https://www.itu.int/rec/T-REC-Y.2060-201206-I.
Jeffery, Keith G. 2021. "FAIR, Open, and Free Does Not Mean No Restrictions." *Patterns* 2 (9).
Jonquet, Clement, John Graybeal, Syphax Bouazzouni, Michael Dorf, Nicola Fiore, Xeni Kechagioglou, Timothy Redmond, et al. 2023. "Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology." In *International Semantic Web Conference*, 38–58. Cham: Springer Nature Switzerland.
Kloog, Itai, Lara Ifat Kaufman, and Kees De Hoogh. 2018. "Using Open Street Map Data in Environmental Exposure Assessment Studies: Eastern Massachusetts, Bern Region, and South Israel as a Case Study." *International Journal of Environmental Research and Public Health* 15 (11): 2443. https://doi.org/10.3390/ijerph15112443.
Kocaman, Sultan, Sameer Saran, Murat Durmaz, and Senthil Kumar. 2021. "Editorial on the Citizen Science and Geospatial Capacity Building." *ISPRS International Journal of Geo-Information* 10 (11): 741. https://doi.org/10.3390/ijgi10110741.
König, Ariane, Karl Pickar, Jacek Stankiewicz, and Kristina Hondrila. 2021. "Can Citizen Science Complement Official Data Sources That Serve as Evidence-Base for Policies and Practice to Improve Water Quality?." *Statistical Journal of the IAOS* 37 (1): 189–204. https://doi.org/10.3233/SJI-200737.
Leadbetter, Adam M. 2015. "Linked Ocean Data." In *The Semantic Web in Earth and Space Science. Current Status and Future Directions*, 11–31. IOS Press.
Liu, Hai-Ying, Daniel Dörler, Florian Heigl, and Sonja Grossberndt. 2021. "Citizen Science Platforms." *The Science of Citizen Science* 22: 439–459.
Magagna, Barbara, Ilaria Rosati, Maria Stoica, Sirko Schindler, Gwenaelle Moncoiffe, Anusuriya Devaraju, Johannes Peterseil, and Robert Huber. 2021. "The I-ADOPT Interoperability Framework for FAIRer Data Descriptions of Biodiversity." arXiv preprint arXiv:2107.06547.
Mankins, John C. 1995. "Technology readiness levels." White Paper, April 6, no. 1995.
Marchezini, Victor, Rachel Trajber, Débora Olivato, Viviana Aguilar Munoz, Fernando de Oliveira Pereira, and Andréa Eliza Oliveira Luz. 2017. "Participatory Early Warning Systems: Youth, Citizen Science, and Intergenerational Dialogues on Disaster Risk Reduction in Brazil." *International Journal of Disaster Risk Science* 8 (4): 390–401. https://doi.org/10.1007/s13753-017-0150-9.
Mons, Barend, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da, Silva Santos, and Mark D. Wilkinson. 2017. "Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud." *Information Services & use* 37 (1): 49–56. https://doi.org/10.3233/ISU-170824.
Mooney, Peter, and Marco Minghini. 2017. "A Review of OpenStreetMap Data." In *Mapping and the Citizen Sensor*, edited by Giles Foody, Linda See, Steffen Fritz, Peter Mooney, Ana-Maria Olteanu-Raimond, Cidália Costa Fonte, and Vyron Antoniou, 37–59. London, UK: Ubiquity Press.
Neis, Pascal, and Dennis Zielstra. 2014. "Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap." *Future Internet* 6 (1): 76–106. https://doi.org/10.3390/fi6010076.
OGC. 2021. "OGC SensorThings API Part 1: Sensing Version 1.1." Accessed August 30, 2023. https://docs.ogc.org/is/18-088/18-088.html#fig-sensing-entities.
OGC. 2022. "Best Practice for using SensorThings API with Citizen Science." Accessed August 30, 2023. https://docs.ogc.org/bp/21-068.pdf.
OSM. 2022. "WikiProject CircularEconomy." Accessed August 30, 2023. https://wiki.openstreetmap.org/wiki/WikiProject_CircularEconomy.
Peng, G. 2023. "Finding Harmony in FAIRness." Eos. Accessed August 30, 2023. https://eos.org/opinions/finding-harmony-in-fairness.
PlanIT, G. E. O. 2023. "A Better Way to Map and Manage Your Urban Forest." Accessed August 30, 2023. https://marketing.planitgeo.com/treeplotter-software-uk.
Poveda-Villalón, María, Paola Espinoza-Arias, Daniel Garijo, and Oscar Corcho. 2020. "Coming to Terms with FAIR Ontologies." In *International Conference on Knowledge Engineering and Knowledge Management*, 255–270. Cham: Springer International Publishing.
Ramachandran, Rahul, Kaylin Bugbee, and Kevin Murphy. 2021. "From Open Data to Open Science." *Earth and Space Science* 8 (5): e2020E–A001562.
Ramírez-Andreotta, Mónica D., Ramona Walls, Ken Youens-Clark, Kai Blumberg, Katherine E. Isaacs, Dorsey Kaufmann, and Raina M. Maier. 2021. "Alleviating Environmental Health Disparities Through Community Science and Data Integration." *Frontiers in Sustainable Food Systems* 5: 620470. https://doi.org/10.3389/fsufs.2021.620470.
Roger, Erin, Dax Kellie, Cameron Slatyer, Peter Brenton, Olivia Torresan, Elycia Wallis, and Andre Zerger. 2023. Open Access Research Infrastructures are Critical for Improving the Accessibility and Utility of Citizen Science: A Case Study of Australia's National Biodiversity Infrastructure, the Atlas of Living Australia (ALA).

Schade, Sven, Marina Manzoni-Brusati, Chrysi Tsinaraki, Alexander Kotsev, K. Fullerton, Roberto Sgnaolin, Fabiano Spinelli, and Irena Mitton. 2017. *Using New Data Sources for Policymaking.* Luxembourg: Publications Office of the European Union.

Schade, Sven, and Chrysi Tsinaraki. 2016. *Survey Report: Data Management in Citizen Science Projects.* Luxembourg: Publication Office of the European Union.

Schentz, Herbert, Johannes Peterseil, and Nic Bertrand. 2013. "EnvThes–Interlinked Thesaurus for Long Term Ecological Research, Monitoring, and Experiments." In *Proceedings of the 27th Conference on Environmental Informatics-Informatics for Environmental Protection, Sustainable Development and Risk Management*. City: Shaker Verlag.

See, Linda. 2019. "A Review of Citizen Science and Crowdsourcing in Applications of Pluvial Flooding." *Frontiers in Earth Science* 7: 44. https://doi.org/10.3389/feart.2019.00044.

See, Linda, Peter Mooney, Giles Foody, Lucy Bastin, Alexis Comber, Jacinto Estima, Steffen Fritz, et al. 2016. "Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information." *ISPRS International Journal of Geo-Information* 5 (5): 55. https://doi.org/10.3390/ijgi5050055.

Shirk, Jennifer L., Heidi L. Ballard, Candie C. Wilderman, Tina Phillips, Andrea Wiggins, Rebecca Jordan, Ellen McCallie, et al. 2012. "Public Participation in Scientific Research: A Framework for Deliberate Design." *Ecology and Society* 17 (2).

Snell, Katherine RS, Rie BE Jensen, Troels E. Ortvad, Mikkel Willemoes, and Kasper Thorup. 2020. "Multiple Fragmented Habitat-Patch Use in an Urban Breeding Passerine, the Short-Toed Treecreeper." *PLoS One* 15 (1): e0227731. https://doi.org/10.1371/journal.pone.0227731.

Spasiano, Andrea, Salvatore Grimaldi, Alessio Maria Braccini, and Fernando Nardi. 2021. "Towards a Transdisciplinary Theoretical Framework of Citizen Science: Results from a Meta-Review Analysis." *Sustainability* 13 (14): 7904. https://doi.org/10.3390/su13147904.

Spear, Dakota M., Gregory B. Pauly, and Kristine Kaiser. 2017. "Citizen Science as a Tool for Augmenting Museum Collection Data from Urban Areas." *Frontiers in Ecology and Evolution* 86.

Stevenson, Robert, Carl Merrill, and Peter Burn. 2021. "Useful Biodiversity Data Were Obtained by Novice Observers Using INaturalist During College Orientation Retreats." *Citizen Science: Theory and Practice* 6 (1).

Sullivan, Brian L., Jocelyn L. Aycrigg, Jessie H. Barry, Rick E. Bonney, Nicholas Bruns, Caren B. Cooper, Theo Damoulas, et al. 2014. The EBird Enterprise: An Integrated Approach to Development and Application of Citizen Science." *Biological Conservation* 169: 31–40. https://doi.org/10.1016/j.biocon.2013.11.003.

Sy, Bocar, Corine Frischknecht, Hy Dao, David Consuegra, and Gregory Giuliani. 2020. "Reconstituting Past Flood Events: The Contribution of Citizen Science." *Hydrology and Earth System Sciences* 24 (1): 61–74. https://doi.org/10.5194/hess-24-61-2020.

Thomas, Usländer, Yves Coene, and Pier Giorgio Marchetti. 2012. "Heterogenous Missions Accessibility." Accessed August 30, 2023. https://www.researchgate.net/publication/258644058_Heterogenous_Missions_Accessibility.

Tom-Aba, Daniel, Adeniyi Olaleye, Adebola Tolulope Olayinka, Patrick Nguku, Ndadilnasiya Waziri, Peter Adewuyi, Olawunmi Adeoye, et al. 2015. "Innovative Technological Approach to Ebola Virus Disease Outbreak Response in Nigeria Using the Open Data kit and Form Hub Technology." *PLoS One* 10 (6): e0131000.

Turicchia, Eva, Massimo Ponti, Gianfranco Rossi, and Carlo Cerrano. 2021a. "The Reef Check Med Dataset on Key Mediterranean Marine Species 2001–2020." *Frontiers in Marine Science* 8: 675574. https://doi.org/10.3389/fmars.2021.675574.

Turicchia, Eva, Massimo Ponti, Gianfranco Rossi, Martina Milanese, Cristina Gioia Di Camillo, and Carlo Cerrano. 2021b. "The Reef Check Mediterranean Underwater Coastal Environment Monitoring Protocol." *Frontiers in Marine Science* 8: 620368. https://doi.org/10.3389/fmars.2021.620368.

Usländer, Thomas, Yves Coene, and Pier G. Marchetti. 2012. "Heterogeneous missions accessibility." *ESA Training Manual*.

Van den Berg, Wouter. 2023. "IDSA Tech Talk | Semantic interoperability." Accessed August 30, 2023. https://www.semantic-treehouse.nl/blog/idsa-tech-talk-semantic-interoperability/.

Wandersman, Abraham. 2003. "Community Science: Bridging the Gap Between Science and Practice with Community-Centered Models." *American Journal of Community Psychology* 31 (3-4): 227–242. https://doi.org/10.1023/A:1023954503247.

Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, and D. Vieglais. 2012. "Darwin Core: An Evolving Community-Developed Biodiversity Data Standard." *PLoS One* 7 (1): e29715. https://doi.org/10.1371/journal.pone.0029715.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 1–9.

Wolff, Erich. 2021. "The Promise of a "People-Centred" Approach to Floods: Types of Participation in the Global Literature of Citizen Science and Community-Based Flood Risk Reduction in the Context of the Sendai Framework." *Progress in Disaster Science* 10: 100171. https://doi.org/10.1016/j.pdisas.2021.100171.

Wolff, Erich, and Felipe Muñoz. 2021. "The Techno-Politics of Crowdsourced Disaster Data in the Smart City." *Frontiers in Sustainable Cities* 3: 695329. https://doi.org/10.3389/frsc.2021.695329.

Appendices

Appendix 1

Table A1. Citizen science information resources and training materials. Note: topics listed here are extracted from the descriptions of the resources.

Citizen Science Information Resources and Training Materials			
Resource / Description	Domain / Audience	Source / Mian topics	Notes
FAIR Data in Citizen Science Projects https://eu-citizen.science/resource/159 A set of resources for research data management and FAIR principles in citizen science.	Any /Librarians Project leaders	EU-Citizen.Science / FAIR data	While designed for librarians it is also applicable to project managers.
Doing Citizen Science as Open Science https://moodle.eu-citizen.science/mod/page/view.php?id=690 A training course that introduces Open Science and covers good practices of data sharing and reuse (including FAIR), ethical aspects, licences, open software and hardware, and more.	Any / Project leaders	EU-Citizen.Science / FAIR data Open Science Open software and hardware CitSci ethics Data sharing Data reuse Data licences	Provides a link to the ***Global Open Science Hardware*** community: https://openhardware.science/.
Basic Regulations and Ethics for Citizen Science https://moodle.eu-citizen.science/course/view.php?id=33 A training course aimed at citizen scientists or volunteers covering their legal status in research projects, rights and obligations, and ethics of gathering and analysing scientific data.	Any / Citizen scientists Volunteers	EU-Citizen.Science / Volunteer legal status Volunteer rights Volunteer obligations Ethics of data gathering Ethics of data analysis Data aggregation Co-authoring	The course comprises information videos and video transcripts. Particular focus is given to the handling of sensitive data (e.g. medical or personal). The course also discusses co-authoring of research papers with citizen scientists.
UK Environmental Observation Network (UKEOF) Resources https://www.ukeof.org.uk/resources/citizen-science-resources Provides resources for environmental citizen science.	Environment / Project leaders	UKEOF and EU-Citizen.Science / Project participation Data management plan Data documentation Meta(data) standards Controlled vocabularies	Data Management Planning resource emphasises the importance of using data and metadata standards and controlled vocabularies to enhance data interoperability and sharing.
Data Ethics for Practitioners https://scistarter.org/training-dataethics A training course that is designed for project leaders and covers ethical principles of collecting data via participatory CitSci.	Any / Project leaders	SciStarter / Ethical principles Ethics of collecting CitSci data Ethics of managing CitSci data Data integrity Data governance Data sharing	The course highlights such topics as data sharing with the scientific community and with the project participants.

Appendix 2

Table A2. Tools that can support citizen science project lifecycle.

Tools that can support citizen science project lifecycle.				
Tool / Description	Governing Body /Funding	Domain	Functions	Notes
Project Discovery Platforms				
EU-Citizen.Science https://eu-citizen.science Citizen science project discovery platform; established in 2019. Lists 268 (218 active) projects. Lists 24 training courses and 802 activities.	Initially funded by the European Union's Horizon programme. Supported by a consortium of 14 partners and 9 third parties from across 14 European countries.	Any /Europe	**Main:** Project discovery **Additional:** Database of tools Training courses Events Forum	Offers a Swagger API to search for and retrieve full projects' metadata.
SciStarter https://scistarter.org/ Online citizen science project discovery platform; founded in 2011. Lists 1528 projects.	Supported by grants from the National Science Foundation, the Institute for Museum and Library Services, Schmidt Futures, NASA, and the National Library of Medicine. Operates as a collaborative effort of academic institutions, federal agencies, non-profit organisations, private foundations, and individual project leaders.	Any /Global	**Main:** Project discovery **Additional:** Project hosting Participation tracking Data visualisation on map Training courses	Offers an API to search for and retrieve full projects' metadata.
CitizenScience.gov https://www.citizenscience.gov US government crowdsourcing and CitSci platform for the US federal government and nongovernmental organisations. Lists 502 projects.	Managed by U.S. General Services Administration.	Any /United States	Main: Project discovery Additional: Information resources	Catalogue of US federally supported projects.
Project hosting				
Anecdata https://anecdata.org Free online citizen science project hosting platform Offers tools to facilitate the entire CitSci project process. Established in 2014. Lists over 300 active projects, 15,500 users, 111,000 observations, and 74,000 photos and mages.	Founded by the Community Lab at the MDI Biological Laboratory in Bar Harbor, Maine.	Any	**Main:** Project hosting **Additional:** Project discovery (limited) Custom data sheets Data collection Mobile app (Android, iOS) Participation management Data sharing Access management Data visualisation on map	It is well suited for more complex biodiversity protocols such as recording absence data, water quality monitoring, litter recording and clean up, and collection of non-biodiversity image observations. Integrates with SciStarter Project and observation data search are limited. Not compliant to FAIR.

(*Continued*)

Table A2. Continued.

Tools that can support citizen science project lifecycle.				
Tool / Description	Governing Body /Funding	Domain	Functions	Notes
CitSci www.citsci.org Offers tools to facilitate the entire CitSci project process. Established in 2007. Lists 1,246 projects, 1,819,943 measurements, 82,904 locations, and 147,504 observations.	The platform is supported by the Natural Resource Ecology Lab (NREL) at Colorado State University.	Any	**Main:** Project hosting **Additional:** Project discovery (limited) Custom data sheets Data collection Mobile app (Android, iOS) Participation management Data analysis Data sharing Access management Data visualisation on map Feedback API for project owners	Projects are automatically added to SciStarter Project search is very limited. Not compliant to FAIR.
CS Center Zürich https://www.citizenscience.uzh.ch/de.html	A joint initiative by the University of Zurich and ETH Zurich, supported by the Mercator Foundation Switzerland	Any	**Main:** Project hosting **Additional:** Training courses	No data collection functionality. The platform only allows to provide information about the project.
FieldScope https://www.fieldscope.org/ Commercial citizen science project hosting platform	Owned and managed by BSCS Science Learning.	Any	**Main:** Project hosting **Additional:** Data collection Map-based data visualisation Analysis of trends, patterns, and changes over time	In addition to hosting citizen science projects, FieldScope supports ***Invitations to Inquiry*** initiative where middle and high school students are offered short learning activities to collect, visualise, and analyse large sets of environmental data.
I See Change https://www.iseechange.com Launched in 2012.	Managed by ISeeChange commercial company.	Climate	**Main:** Project hosting **Additional:** Data collection Mobile app (Android, iOS)	
Just One Giant Lab (JOGL) https://app.jogl.io/search/projects	Operated by Just One Giant Lab. Funded by voluntary donations.	Any	**Main:** Project hosting	

nQuire https://nquire.org.uk/explore Launched in 2017.	Operated by the Institute of Educational Technology, The Open University		**Additional:** Project discovery **Main:** Project hosting **Additional:** Project discovery Data collection	
SciFabric https://scifabric.com/projects/ Launched in 2015.	Operated by Scifabric. Supported by Shuttleworth Foundation.	Any	**Main:** Project hosting **Additional:** –	Scifabric offers paid subscriptions for crowdsourcing projects. Uses PYBOSSA open-source crowdsourcing framework.
Scotland's environment Citizen Science Portal https://envscot-csportal.org.uk/	Managed by the Scottish Environment Protection Agency (SEPA).	Environment	**Main:** Project hosting **Additional:** Data collection Mobile app (Android, iOS) Data visualisation on map	Projects are hosted by SEPA for public contribution.
World Community Grid https://www.worldcommunitygrid.org/research/projects.s	Operated by World Community Grid.	Any	**Main:** Project hosting **Additional:** Project discovery (limited) Data collection	Requires BOINC software application to contribute data.
Zooniverse www.zooniverse.org Crowdsourcing CitSci platform for classification, identification, transcription tasks. Launched in 2007 Lists 97 active, 243 paused, and 110 finished projects.	Owned and operated by the Citizen Science Alliance.	Any	**Main:** Project hosting **Additional:** Project discovery (limited) Mobile app (Android, iOS) Participation tutorials Participation management Custom workflows (sequence of tasks) Access management	Completed projects can publish aggregated results and reports.
Data Collection				
ALA BioCollect https://www.ala.org.au/biocollect/ Established in 2010.	Operated by Atlas of Living Australia.	Biodiversity	**Main:** Data collection **Additional:** Data storage	The ALA receives support from the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) and is hosted by the

Table A2. Continued.

Tools that can support citizen science project lifecycle.				
Tool / Description	Governing Body /Funding	Domain	Functions	Notes
				Commonwealth Scientific and Industrial Research Organisation (CSIRO).
ArcGIS Survey123 https://survey123.arcgis.com/ Fully customisable survey product for data collection via a web-browser or a mobile app.	Managed by Esri commercial company.	Any	**Main:** Data collection **Additional:** Data storage	Commercial product. Needs to be purchased as part of ArcGIS Online toolkit.
ArcGIS QuickCapture https://www.esri.com/en-us/arcgis/products/arcgis-quickcapture/overview A survey product for field observations from the devices on moving vehicles.	Managed by Esri commercial company.	Any	**Main:** Data collection **Additional:** Data storage	Commercial product. Needs to be purchased as part of ArcGIS Online toolkit.
ArcGIS Community Science Solution https://doc.arcgis.com/en/arcgis-solutions/11.0/reference/introduction-to-community-science.htm A product specifically designed for collecting location-enabled plant and animal observations from citizen scientists.	Managed by Esri commercial company.	Biodiversity	**Main:** Data collection **Additional:** Data storage	Commercial product. Specifically designed for citizen science. Needs to be purchased as part of ArcGIS Online toolkit.
conserve.io https://conserve.io/	Operated by Conserve.IO.	Biodiversity	**Main:** Data collection **Additional:** Mobile app (Android, iOS)	Supported projects: Whale Alert Sharktivity Manatee Alert
CyberTracker https://cybertracker.org/ CyberTracker Icon User Interface was developed in 1996.	Operated by CyberTracker Conservation.	Biodiversity	**Main:** Data collection **Additional:** Mobile app (Android, iOS) Training	CyberTracker Online provides a free software service, simplified for nontechnical users who cannot afford expensive technical support. The CyberTracker Icon User Interface was designed for expert indigenous Kalahari San Master Trackers who cannot read or write.
eBird https://ebird.org/home Established 2002.	Operated by Cornell Lab.	Biodiversity	**Main:** Data collection **Additional:** Mobile app (Android, iOS) Data sharing	Users can install eBird app and contribute bird sightings to the eBird platform.
eButterfly https://www.e-butterfly.org/ Launched in 2011.	Managed by a team of volunteers.	Biodiversity	**Main:** Data collection **Additional:** Project hosting Project discovery (limited) Data storage Data sharing	Lists over 1,500 species and over 500,000 observations.

EpiCollect 5 https://five.epicollect.net/ Open source data collection platform	Managed by the Centre for Genomic Pathogen Surveillance (CGPS) and the commercial arm Digital Epidemiology Services LTD (DES)	Any	**Main:** Data collection **Additional:** Project creation Project hosting Custom datasheets Project discovery (limited) Mobile app (Android, iOS) Data storage Data sharing Data visualisation on map Data download in CSV and JSON format	Primary function is mobile data-gathering in a broad sense, rather than being exclusively tailored to citizen science initiatives. Media uploads are not limited but Epicollect5 is not a data storage platform.
Global Geo-Referenced Field Photo Library https://www.ceom.ou.edu/photos/	Managed by Center for Earth Observation and Modeling, the University of Oklahoma.	Earth Observations	**Main:** Data collection **Additional:** Data visualisation on map	Citizen science project that collects geo-referenced photos in the fields.
GIS Cloud https://www.giscloud.com/ Launched in 2010.	Managed by GIS Cloud Ltd.	Any	**Main:** Data collection **Additional:** Custom datasheets Mobile app (Android, iOS) Data storage Data visualisation on map Data analysis	Commercial real-time mapping software.
GLOBE Observer https://observer.globe.gov/ the GLOBE Program founded in 1995.	GLOBE Observer is funded under the NASA Earth Science Education Collaborative (NESEC).	Cloud Cover Mosquito Habitats Land Cover Trees Eclipses	**Main:** Data collection **Additional:** Mobile app (Android, iOS) Data sharing	Users can install GLOBE Observer app and contribute data to the GLOBE platform.
Google forms https://www.google.com/forms/about/	Operated by Google.	Any	**Main:** Data collection **Additional:** Custom datasheets Data storage	Google Forms is a cloud-based survey software included as part of the free, web-based Google Docs Editors suite offered by Google.
iMoustique https://www.eidatlantique.eu/page.php?P=155 Launched in 2013.	Developed and maintained by EID Atlantique.	Biodiversity	**Main:** Data collection **Additional:** Mobile app (Android, iOS) Training material	A mobile app for taking photos of mosquitoes and indicating the presence of mosquitoes, including the 'tiger mosquito'.
	iNaturalist became an independent nonprofit organization in July 2023.	Biodiversity	**Main:** Data collection **Additional:** Mobile	Users can install iNaturalist app and contribute data to the iNaturalist platform.

Table A2. Continued.

Tools that can support citizen science project lifecycle.				
Tool / Description	Governing Body /Funding	Domain	Functions	Notes
iNaturalist https://www.inaturalist.org/ Founded in 2008.			app (Android, iOS) Data sharing	The platform employs image recognition technology and artificial intelligence algorithms to suggest identifications for the uploaded observations.
Ispot https://www.ispotnature.org/ Launched in 2009.	Received funding from the Garfield Weston Foundation, the British Ecological Society and is part of The OpenScience Laboratory, an initiative of The Open University and The Wolfson Foundation.	Biodiversity	**Main:** Data collection **Additional:** Species identification	Registered users can contribute observations and engage in species identification.
Open Data Kit (ODK) https://getodk.org/ Established in 2008.	Managed by ODK Inc. Funded via ODK Cloud subscription plans.	Any	**Main:** Data collection **Additional:** Data storage	Open-source data collection toolkit for mobile devices. Support efficient and reliable data collection in offline or low-connectivity. Additionally offers ODK Cloud – a fully managed subscription-based platform.
ODK-X https://odk-x.org/	Managed by ODK Inc. Funded via ODK Cloud subscription plans.	Any	**Main:** Data collection **Additional:** Data storage	Free and open-source software for collecting, managing, and using data in resource-constrained environments. More complex version of ODK.
OSM Tracker https://learnosm.org/en/mobile-mapping/osmtracker/	Managed by HOT OSM.	Any	**Main:** Data collection **Additional:** Mobile app (Android, iOS, Windows)	OSMTracker allows the creation of a gpx trace of a journey, with the collection of 'waypoints' along the route. Voice recording, photographs and other notes may also be recorded, and all will be geolocated.
PlantNet https://plantnet.org/en/ Launched in 1996.	PlantNet is an open consortium founded by four French research organizations (CIRAD, Inria, INRAE, IRD) and supported by Agropolis Fondation.	Biodiversity	**Main:** Data collection **Additional:** Mobile app (Android, iOS)	Pl@ntNet is a citizen science project available as an app that helps identify plants. Registered users can contribute observations and engage in species identification.
Pybossa https://pybossa.com/ Open-source crowdsourcing and data enhancement server platform.	Managed and financed by SciFabric commercial company.	Any	**Main:** Data enhancement **Additional:** Data collection Data storage	Requires extensive technical skills to install and use the server.
Sapelli https://www.sapelli.org/	Developed as part of an EPSRC funded project called Intelligent Maps by the Extreme Citizen Science (ExCiteS) research group at University College London.	Any	**Main:** Data collection **Additional:** Mobile app (Android)	Open-source project that facilitates data collection across language or literacy barriers through highly configurable icon-driven user interfaces.
SPOTTERON https://www.spotteron.net/	Owned by a commercial company SPOTTERON.	Any	**Main:** Data collection **Additional:** Project	Commercial product.

Commercial web-based platform that facilitates citizen science projects and environmental monitoring initiatives			creation Project hosting Custom datasheets Mobile app (Android, iOS) Data sharing Data quality Customisable project website Community support	
Tigatrapp https://github.com/MoveLab/tigatrapp-android Launched in 2013.	Developed by the Movement Ecology Laboratory of the Center for Advanced Studies of Blanes (CEAB-CSIC)	Biodiversity	**Main:** Data collection **Additional:** Mobile app (Android)	Developed as part of the citizen science project Mosquito Alert.
Ushahidi https://www.ushahidi.com Founded in 2008.	Operated by Ushahidi, Inc.	Any	**Main:** Data collection **Additional:** Mobile app (Android, iOS) Data storage Data sharing Data analysis Data visualisation Data download in CSV	The Ushahidi Platform offers accessible crowdsourcing and mapping tool. Free Basic plan is available for small non-profit and grassroot organisations.
Semantic Resources				
AGROVOC https://agrovoc.fao.org/browse/agrovoc/en/ Established in the early 1980s.	Managed by the Food and Agriculture Organization of the United Nations (FAO)	Agriculture	**Main:** Controlled vocabulary	
Chemical Entities of Biological Interest Ontology (ChEBI) https://www.ebi.ac.uk/chebi/aboutChebiForward.do Established in 2007.	ChEBI was funded by BBSRC.	Chemical sciences	**Main:** Controlled vocabulary	ChEBI is part of the ELIXIR infrastructure
Darwin Core https://dwc.tdwg.org/ A semantic resource and a structural data standard for publishing, integrating and sharing biodiversity information. Ratified as a standard in 2009.	Maintained by the Darwin Core Maintenance Interest Group	Biodiversity	**Main:** Data standard Vocabulary	
Environment Ontology (ENVO) https://sites.google.com/site/environmentontology/ Established in 2013.	Established with the support of ESIP Federation, UN Environment, IOC-UNESCO, and individual contributions.	Environmental	**Main:** Controlled vocabulary	The ontology can be downloaded from OntoBee web server, EBI Ontology Lookup Service repository, or EnvO GitHub repository.
	Established and managed by INSPIRE.	Geospatial		Groupings of spatial data themes.

Table A2. Continued.

Tools that can support citizen science project lifecycle.				
Tool / Description	Governing Body /Funding	Domain	Functions	Notes
GEMET/INSPIRE http://vocab.nerc.ac.uk/collection/P22/current/			**Main:** Controlled vocabulary	
The NERC Vocabulary Server (NVS) https://vocab.nerc.ac.uk/ A collection of standardised and hierarchically structured controlled vocabularies	Maintained by the National Oceanography Centre – British Oceanographic Data Centre (BODC).	Oceanographic and related domains	**Main:** Controlled vocabulary	An Interactive Query UI tool provides a simple interface to query NVS triplestore (RDF database of all NVS vocabularies) using SPARQL queries
The Exposure Ontology https://bioportal.bioontology.org/ontologies/EXO Established in 2022.	Development of the ontology was supported by the American Chemistry Council Long Range Research Initiative.	Environment	**Main:** Controlled vocabulary	Designed for exposure science and diverse environmental health disciplines including toxicology, epidemiology, disease surveillance, and epigenetics.
University of Arizona Superfund Research Project Data Interface Ontology (SRPDIO) https://github.com/UA-SRC-data/srpdio	This project was funded by the National Institute of Environmental Health's Superfund Research Program.	Environment	**Main:** Controlled vocabulary	SRPDIO uses descriptive terms from the Environment Ontology (ENVO), the Plant Ontology (PO) and the Phenotype and Trait Ontology (PATO), along with the observation model from the BioCollections Ontology (BCO), to merge quantitative data into a single searchable database.
Data Publishing and Preservation				
Mendeley Data https://data.mendeley.com/ Launched in 2016.	A product of Elsevier.	Any	**Main:** Data repository	Offers versioning of submissions. Only accessible to registered research institutions.
Zenodo https://zenodo.org/ A multi-disciplinary open repository designed for research communities. Launched in 2013.	Owned by the European Organization for Nuclear Research (CERN).	Any	**Main:** Data repository	Offers versioning of submissions. Offers RESTful API to support deposit of research outputs, records search, and files upload and download.
Standards				
Access to Biological Collection Data (ABCD) https://www.tdwg.org/standards/abcd/	Published by Biodiversity Information Standards (TDWG)	Biodiversity	**Main:** Data standard	
ISO 19115 https://www.iso.org/standard/53798.html	Managed by International Organization for Standardization (ISO)	Geographic information	**Main:** Metadata standard	Defines the schema required for describing geographic information and services by means of metadata.
OGC O&M / ISO 19156:2011 http://www.opengis.net/doc/as/om/2.0 An international standard that defines a conceptual framework and encoding for describing observations and measurements.	Maintained by the Open Geospatial Consortium.	Geographic information systems	**Main:** Data standard	The core of OGC Sensor Web Enablement (SWE) standards such as SensorThings API, WaterML 2.0, and Sensor Observation Service (SOS).
OGC SensorThings API 1.1 (STA) https://docs.ogc.org/is/18-088/18-088.html#fig-	Maintained by the Open Geospatial Consortium.	Any	**Main:** Data standard	

sensing-entities An international standard for interconnecting heterogeneous Internet of Things (IoT) devices. The first version published in 2016.				
OGC STAplus 1.0 https://docs.ogc.org/DRAFTS/22-022.html An international standard and is an extension to the STA data model that is based on requirements from the citizen science community. Approved as a standard in 2023.	Maintained by the Open Geospatial Consortium.	Any	**Main:** Data standard	Specific to citizen science.
PPSR Core https://core.citizenscience.org/ An open data and metadata standard that defines a common framework for describing citizen science projects.	Maintained by the Citizen Science Association's Data & Metadata Working Group	Any	**Main:** Data standard Metadata standard	Specific to citizen science.
Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR) https://www.ospar.org/aboutDeveloped from the 1972 Oslo Convention on dumping waste at sea and the 1974 Paris Convention on land-based sources of marine pollution.	Managed by the OSPAR Commission.	Marine Environment	**Main:** Convention	A legislative instrument that regulates international cooperation on environmental protection in the North-East Atlantic region. Aims to protect the marine environment from various threats and covers areas such as pollution, biodiversity, and ecosystems.
Quality Assurance / Quality Control of Real Time Oceanographic Data (QARTOD) https://cdn.ioos.noaa.gov/media/2020/10/Prospects-for-RT-QC-manual_5.2.pdf Established in 2003.	Managed by The United States Integrated Ocean Observing System.	Marine Environment	**Main:** Data standard	Quality control steps for documenting the reliability of the collected real-time environmental data.
Ecological Metadata Language (EML) https://eml.ecoinformatics.org/ Version 2.1.1 released in 2011.	The specification is maintained by voluntary project members.	Ecology	**Main:** Metadata standard	
Data Catalog Vocabulary (DCAT) https://www.w3.org/TR/vocab-dcat-3/ Standardised in 2014.	Standardised by the Government Linked Data (GLD) Working Group.	Any	**Main:** Metadata standard	DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web.