Project title:	All Data 4 Green Deal - An Integrated, FAIR Approach for the Common European Data Space
Project number:	101061001
Project Acronym:	AD4GD
Type:	HORIZON-AG - HORIZON Action Grant Budget-Based
Work program topics addressed:	HORIZON-CL6-2021-GOVERNANCE-01

DELIVERABLE NO: D2.3

CITSCI DATA INTEROPERABILITY OBSTACLES AND SOLUTIONS

Due date of deliverable:	31/08/2023
Actual submission date:	18/10/2023
Version:	1.0
Main Authors:	Victoria Lush, Lucy Bastin, Joan Masó

Co-funded by the European Union, Switzerland and the United Kingdom

Document Metadata

Project number	101061001
Project title	All Data 4 Green Deal - An Integrated, FAIR Approach for the Common European Data Space

Deliverable title	CitSci data interoperability obstacles and solutions
Deliverable number	D2.3
Deliverable version	1
Contractual date of delivery	31/08/2023
Actual date of delivery	18/10/2023
Document status	Final
Document version	1.0
Online access	URL
Dissemination	Public
Work package	WP2: In situ networks, CitSci and Socioeconomic Data
Partner responsible	Aston University (AU)
Author(s)	Victoria Lush, Lucy Bastin, Joan Masó
Editor(s)	
Reviewer(s)	Francesca Noardo
EC Project Officer	Lara Congiu

Abstract	As part of the European Data Strategy towards establishing a Single EU Market for data, 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 volunteered geographic information and citizen science data. 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. 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". This document is the Deliverable 2.3 for the AD4GD project. It presents the results achieved in the context of Task 2.2 "Identification of CitSci data including complementary socioeconomic and INSPIRE data" as part of Work Package 2 "In-situ networks, CitSci and Socioeconomic Data". The reader should consider that a follow-up version of this document will be released on month 30 of the project as Deliverable 2.4. This document offers a review of the current state of play with citizen science data in regard to FAIR compliance. It highlights citizen science initiatives that successfully apply open standards and tools to collect and share data that

Co-funded by the European Union, Switzerland and the United Kingdom

	closely align with FAIR data principles and is a potential candidate for the inclusion into the EU GDDS. It examines platforms and tools that can support citizen science projects and discusses gaps that still exist to achieve citizen science data FAIRness.
Keywords	Citizen Science, FAIR data principles, In-situ, Metadata standards, Data standards, Tools, APIs.
Disclaimer	Views and opinions expressed in this deliverable are those of the author(s) only and do not necessarily reflect those of the European Union the United Kingdom or Switzerland. Neither the European Union nor United Kingdom nor Switzerland can be held responsible for them

Co-funded by the European Union, Switzerland and the United Kingdom

DOCUMENT VERSION HISTORY

Version history			
Version	**Date**	**Modification reason**	**Modified by**
0.1	06/09/2023	Final draft of the document	Victoria Lush Lucy Bastin Joan Masó
0.2	12/09/2023	Review of the final draft of the document	Francesca Noardo
0.3	20/09/2023	Comments from the review addressed	Victoria Lush
0.3	12/10/2023	Review of the updated draft of the document	Joan Masó
0.4	16/10/2023	Comments from the review addressed	Victoria Lush
0.5	16/10/2023	Review of the updated draft of the document	Lucy Bastin
1.0	17/10/2023	Final version of the document	Victoria Lush

Co-funded by the European Union, Switzerland and the United Kingdom

ABBREVIATIONS

Abbreviation	Definition
ABCD	Access to Biological Collections Data
ALA	Atlas of Living Australia
API	Application Programming Interface
CAP	CERN Analysis Preservation infrastructure
CDM	Common Data Model
CERN	European Organization for Nuclear Research
CGPS	Centre for Genomic Pathogen Surveillance
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DEIMS	Dynamic Ecological Information Management System
DES	Digital Epidemiology Services LTD
DMM	Dataset Metadata Model
DOI	Digital Object Identifier
DwC-A	Darwin Core Archive
EC	European Commission
ECSA	European Citizen Science Association
Esri	Environmental Systems Research Institute
eLTER	European Long-Term Ecosystem, critical zone and socio-ecological Research
EnvThes	Environmental Thesaurus
EU	European Union
FAIR	Findable Accessible Interoperable Reusable
GBIF	Global Biodiversity Information Facility
GDDS	Green Deal Data Space
GLOBE	Global Learning and Observations to Benefit the Environment
GPS	Global Positioning System
HTML	Hypertext Markup Language
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
ODbL	Open Data Commons Open Database License
ODK	Open Data Kit
ODM	Observation Data Model
OGC	Open Geospatial Consortium
IoT	Internet of Things
IPT	Integrated Publishing Toolkit

Co-funded by the European Union, Switzerland and the United Kingdom

IUCN	International Union for Conservation of Nature
NASA	National Aeronautics and Space Administration
NCRIS	National Collaborative Research Infrastructure Strategy
NERC	Natural Environment Research Council
NOC	National Oceanography Centre
NREL	Natural Resource Ecology Lab
NVS	The NERC Vocabulary Server
O&M	Observations and Measurements
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OpenAIRE	Open Access Infrastructure for Research in Europe
ORCID	Open Researcher and Contributor ID
OSM	OpenStreetMap
PANGEA	Publishing Network for Geoscientific and Environmental Data
PMM	Project Metadata Model
PPSR	Public Participation in Scientific Research
R&I	Research and Innovation
RDF	Resource Description Framework
RESTful	Representational State Transfer
SKOS	Simple Knowledge Organization System
STA	SensorThings API
STAplus	SensorThings API plus
UMLS	Unified Medical Language System
XML	Extensible Markup Language

Co-funded by the European Union, Switzerland and the United Kingdom

Table of Contents

Table of Tables

Co-funded by the European Union, Switzerland and the United Kingdom

Table of Figures

Co-funded by the European Union, Switzerland and the United Kingdom

Co-funded by the European Union, Switzerland and the United Kingdom

EXECUTIVE SUMMARY

The common European Green Deal Data Space (GDDS) will interconnect currently fragmented and dispersed data from various sources including private sector, public sector, and citizen science to support the objectives of the European Green Deal. Citizen science data can play an important role in complementing official data sources. While the perception of the quality of such data remains a concern, 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.

To be fit for integration into the Green Deal 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 (Findable, Accessible, Interoperable and Reusable) data principles that emphasise machine-actionability. While FAIR data principles are starting to be promoted by citizen science platforms and initiatives, sources of FAIR citizen science data beyond biodiversity domain are still scarce.

Leading citizen science project discovery platforms (that facilitate search of citizen science projects hosted on the platform itself and/or other sites) such as EU-Citizen.Science might be considered as a potential source of citizen science data, however, at present, these primarily focus on project discovery by prospective participants or collaborators, rather than curation of project data. Some national citizen science networks provide their own project discovery platforms but none of the ones that we analysed support access or links to projects' data. While platforms for discovering citizen science projects are growing, these are unlikely to serve as centralised citizen science data hubs due to lack of necessary technical resources and heterogeneous data licensing issues, since there is no requirement for the projects to provide Open Data. There are several data aggregation facilities (see below) which do permit the collation and discovery of citizen science data, but these are in relatively specific domains (e.g. species presence observations, or data from particular environmental sensors). On the whole, the concept of general data aggregation, curation and discovery is typically considered beyond the intended scope of existing citizen science platforms, and has thus been detected as a gap in this study.

The Global Biodiversity Information Facility (GBIF) is an example data provider that closely aligns with FAIR data principles and offers both official and citizen science data. It can be considered a potential source of biodiversity data for the GDDS. While GBIF data adheres to a well-defined data standard (Darwin Core) that facilitates data interoperability, this also limits the types of data that can be contributed to GBIF. Data that cannot be transformed into the GBIF data model needs to be hosted elsewhere, which creates data fragmentation. OpenStreetMap and Sensor.Community are examples of citizen science initiatives (and data aggregation platforms) that align with some FAIR data principles and can offer valuable inputs to the GDDS. However, an additional layer of applications and semantic resources will be required to facilitate data discovery and data integration with other sources to support the GDDS users. An ongoing research question here is whether the GDDS should specify data models similar to GBIF, or whether the thematic Research Infrastructures that can serve as the intermediaries connecting to the GDDS should be responsible for developing such data models.

There are thousands of active and completed citizen science projects that might hold valuable inputs for the GDDS, but do not adhere to FAIR data principles. To achieve data FAIRness, projects must follow good practices 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 in all stages of the project lifecycle. Adding to the challenge, free and open source tools typically deliver several but not all functions to deliver a project end-to-end solution.

To ensure long-term value outside of 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 often use their own infrastructure which makes data hard to discover (these are mainly focused on data visualisation). Other projects might collect data suited for contribution to larger initiatives that already provide open data

Co-funded by the European Union, Switzerland and the United Kingdom

capabilities, e.g., iNaturalist, eBird, BioCollect, NASA Globe, Sensor.Community. However, such projects might find themselves collecting additional valuable data that does not conform to the accepted data structure of their chosen publishing initiative. 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. Examples of existing open repositories include Zenodo and PANGEA, though both are more suitable for static data snapshots than for the continually-evolving datasets which emerge from many citizen science projects. The inability to support longer-term or ongoing citizen science projects that seek to continually supply their datasets is a major limitation.

Semantics and the use of controlled vocabularies will also play a crucial role in supporting citizen science data FAIRness. To correctly interpret, (re-)use and integrate data, especially across domains, data needs to be properly described, which is not always a straightforward task. Selecting a suitable semantic resource (e.g., a vocabulary) can be challenging because, while many semantic resources exist, there might not be a single resource (or any resources) suited to fully describe the data at hand. Sometimes several vocabularies exist but their concepts are not yet linked, making it currently impossible to automatically derive a match between data described with concepts from those different vocabularies. Semantic resources are essential for data FAIRness – (meta)data standards, controlled vocabularies or other structured data descriptions (e.g., data tagging) facilitate data discovery, sharing, interoperability, (re)use, and integration (especially important across domains). Such resources, and the necessary links between vocabularies, should 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 their own vocabularies from scratch.

International data standards can ensure consistency and interoperability among data collected by different individuals or groups participating in citizen science initiatives. Their use enhances the credibility and scientific value of citizen science efforts, making the data more reliable for researchers, policymakers, and the broader community. Foundational data standards such as Observations and Measurements (O&M) or practical data access protocols such as SensorThings API (STA) can serve as the basis for tailored extensions to meet the needs of citizen science initiatives. Supporting documents similar to '*OGC Best Practice for using SensorThings API with Citizen Science*' will play an important role in providing use cases and in improving understanding of how standards can be applied in practice.

Co-funded by the European Union, Switzerland and the United Kingdom

1 INTRODUCTION

The European Union has adopted the European Green Deal 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, 2023a).

The European Green Deal focuses on eight key themes (EC, 2020) with an ambition to improve well-being and health of citizens and future generations:

1. Increasing climate ambition
2. Clean, affordable and secure energy
3. Industry for a clean and circular economy
4. Energy and resource efficient buildings
5. Sustainable and smart mobility
6. Farm to fork
7. Biodiversity and ecosystems
8. Zero-pollution, toxic-free environments.

As part of the European Data Strategy towards establishing a Single EU Market for data, the European Commission is working on common European data spaces across several domains to enable easy data flow between countries and sectors (EC, 2022). To become viable and useful to decision makers, a set of data spaces is being proposed, including a Green Deal Data Space (GDDS) that covers issues such as climate change, circular economy, pollution, biodiversity and deforestation (Farrell *et al.*, 2023). 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 as a set of guidelines to enhance the value of digital assets (Wilkinson *et al.*, 2016). These principles specifically focus on machine capability to automatically find, access, interoperate, and reuse assets, thereby promoting open data and open science. The FAIR principles are known by the 4 words:

Findable - The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

Accessible - Once the user finds the required data, they need to know how those data can be accessed, possibly including authentication and authorisation.

Interoperable - The data usually needs to be integrated with other data. In addition, the data needs to semantically interoperate with applications or workflows for analysis, storage, and processing.

Reusable - The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings. (GO FAIR, 2016.)

There is no full agreement in the scientific community on how FAIR principles should be evaluated in practice (Peng, 2023). The Go FAIR initiative that works towards implementing the FAIR data principles defines a FAIR assessment framework that consists of 10 principles and sub-principles, with a total of 15 criteria (GO FAIR, 2016), presented in Table 1 below.

Co-funded by the European Union, Switzerland and the United Kingdom

Table 1: Full version of the FAIR principles: Findable (F1, F2, F3, F4), Accessible (A1, A1.1, A1.2, A2), Interoperable (I1, I2, I3), and Reusable (R1, R1.1, R1.2, R1.3).

Principles Name	Principles ID	Principles Description
Findable	F1	(Meta)data are assigned globally unique and persistent identifiers
	F2	Data are described with rich metadata
	F3	Metadata clearly and explicitly include the identifier of the data they describe
	F4	(Meta)data are registered or indexed in a searchable resource
Accessible	A1	(Meta)data are retrievable by their identifier using a standardised communication protocol
	A1.1	The protocol is open, free and universally implementable
	A1.2	The protocol allows for an authentication and authorisation procedure where necessary
	A2	Metadata should be accessible even when the data is no longer available
Interoperable	I1	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
	I2	(Meta)data use vocabularies that follow the FAIR principles
	I3	(Meta)data include qualified references to other (meta)data
Reusable	R1	(Meta)data are richly described with a plurality of accurate and relevant attributes
	R1.1	(Meta)data are released with a clear and accessible data usage licence
	R1.2	(Meta)data are associated with detailed provenance
	R1.3	(Meta)data meet domain-relevant community standards

It is a frequent misconception that FAIR implies Open. Data can be discoverable and accessible via an open protocol, but could still be accompanied with a licence with access conditions and requiring some payment or other compensations. This is important for the development of data spaces that enable the development of a digital economy based on a mixture of open and closed data.

1.1 CITIZEN SCIENCE

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), which sets a goal of bringing science closer to citizens and promoting citizen involvement in scientific processes. The Open Science Policy of the European Commission recognises Citizen Science as one of its eight policy ambitions,

Co-funded by the European Union, Switzerland and the United Kingdom

stating that "the general public should be able to make significant contributions and be recognised as valid European science knowledge producers" (EC, 2019).

Citizen Science can be defined as general public or non-expert participation in scientific processes to produce or enrich scientific knowledge. Citizen science term primarily emerged from the field of biodiversity (Bonney *et al.*, 2009), but it is also known as public participation in scientific research (PPSR) (Shirk *et al.*, 2012), community science (Wandersman, 2003), civic science (Bäckstrand, 2003), amateur science (Gura, 2013), etc. 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 vary, most of them are collecting data as part of their activities.

The European Citizen Science Association (ECSA) outlines '*10 Principles of Citizen Science*' (Table 2) developed by the '*Sharing best practice and building capacity*' Working Group of the ECSA with input from the members of the Association.

Guideline 7 of the principles 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), which implies adherence to data sharing principles as well as FAIR data principles.

Table 2: ECSA 10 Principles of Citizen Science (ECSA, 2015).

Some of the key principles which as a community we believe underlie good practice in citizen science.

1. Citizen science projects actively involve citizens in scientific endeavour that generates new knowledge or understanding. Citizens may act as contributors, collaborators, or as project leader and have a meaningful role in the project.
2. Citizen science projects have a genuine science outcome. For example, answering a research question or informing conservation action, management decisions or environmental policy.
3. Both the professional scientists and the citizen scientists benefit from taking part. Benefits may include the publication of research outputs, learning opportunities, personal enjoyment, social benefits, satisfaction through contributing to scientific evidence e.g., to address local, national and international issues, and through that, the potential to influence policy.
4. Citizen scientists may, if they wish, participate in multiple stages of the scientific process. This may include developing the research question, designing the method, gathering and analysing data, and communicating the results.
5. Citizen scientists receive feedback from the project. For example, how their data are being used and what the research, policy or societal outcomes are.
6. Citizen science is considered a research approach like any other, with limitations and biases that should be considered and controlled for. However unlike traditional research approaches, citizen science provides opportunity for greater public engagement and democratisation of science.
7. Citizen science project data and meta-data are made publicly available and where possible, results are published in an open access format. Data sharing may occur during or after the project, unless there are security or privacy concerns that prevent this.
8. Citizen scientists are acknowledged in project results and publications.
9. Citizen science programmes are evaluated for their scientific output, data quality, participant experience and wider societal or policy impact.
10. The leaders of citizen science projects take into consideration legal and ethical issues surrounding copyright, intellectual property, data sharing agreements, confidentiality, attribution, and the environmental impact of any activities.

Citizen science data can play an important role in complementing official data sources, not least because of their currency and specificity. While the quality of such data is perceived as a concern, the development of consistent study protocols, advanced data collection and data visualisation tools, data standards and

Co-funded by the European Union, Switzerland and the United Kingdom

protocols, machine learning for calibration and outlier detection can help address some data quality issues. Yet, to be fit for integration into the Green Deal Data Space and to 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 outlined above.

In this document, we review the current state of play with citizen science data in regard to FAIR compliance. We first highlight citizen science initiatives that successfully apply open standards and tools to collect and share community-generated data. Then, we examine platforms and tools that can support citizen science projects at different stages of their lifecycle, from project initiation to open data sharing. Finally, we discuss gaps that still exist and consider open standards that could support citizen science data FAIRness.

The document is structured as follows:

- Section 2 discusses citizen science project discovery platforms that facilitate search of citizen science projects.
- Section 3 discusses data aggregation platforms that collect domain-specific data and facilitate data sharing.
- Section 4 examines tools that support citizen science including information resources, project hosting platforms, data collection tools, and semantic resources.
- Section 5 reviews standards for structuring and accessing data.
- Section 6 discusses larger-scale ongoing and completed projects that work(ed) towards creating tools for and supporting citizen science initiatives.
- Section 7 offers conclusions and discusses future work.

2 CITIZEN SCIENCE PROJECT DISCOVERY PLATFORMS

Citizen science project discovery platforms are platforms that facilitate search of citizen science projects hosted on the platform itself and/or other sites. Such platforms might be considered as a potential source of citizen science data, however, at present these primarily focus on project discovery by prospective participants or collaborators, rather than curation of project data. This is not surprising as citizen science projects are always concerned with engaging with new collaborators and participants to maintain the momentum of the activity. Some of the leading citizen science project discovery platforms, those that are not domain-specific, are ***SciStarter***, ***CitizenScience.gov***, and ***EU-Citizen.Science***. We explore these platforms below.

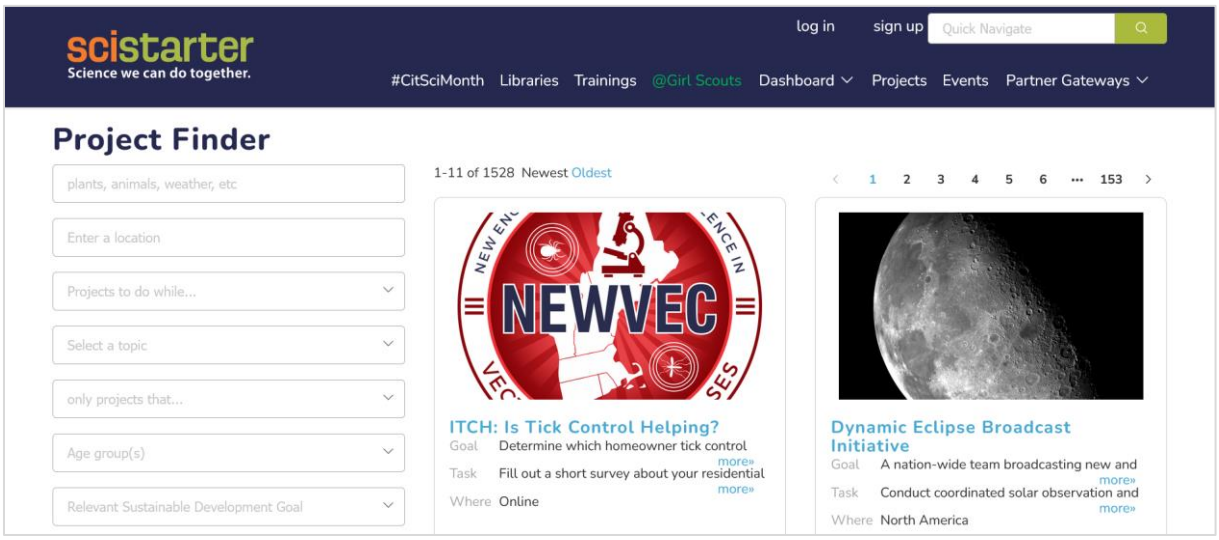


Figure 1: SciStarter platform – Project Finder page.

Co-funded by the European Union, Switzerland and the United Kingdom

SciStarter is a citizen science project discovery platform founded by Darlene Cavalier (University of Pennsylvania, US) in 2011 (SciStarter, 2023a). At present, the platform is primarily supported by grants from the National Science Foundation, the Institute for Museum and Library Services, Schmidt Futures, NASA, and the National Library of Medicine. SciStarter operates as a collaborative effort involving multiple partners, including academic institutions, federal agencies, non-profit organisations, private foundations, and individual project leaders.

SciStarter is a global platform covering any thematic areas and is more popular among the US-based projects. It contains 1528 registered projects, some of which are hosted directly on SciStarter platform (e.g., see Figure 2). The platform offers data hosting which allows users to submit their observations and permits visualisation of observations on a map (e.g., see Figure 3). This enables potential re-users to better evaluate project data for fitness-for-use, but raw observation data is not accessible to download.

Figure 2: SciStarter platform – ZomBee Watch hosted project page.

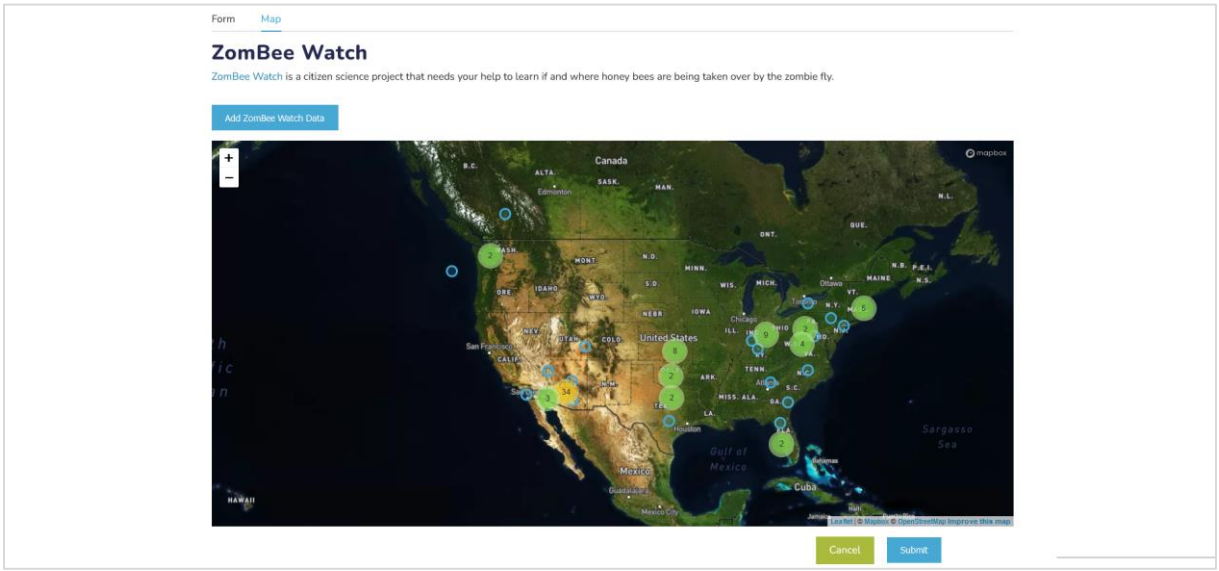


Figure 3: SciStarter platform – ZomBee Watch observations map.

The Project Finder allows searching for projects by keywords, location, activity (e.g., while on a hike, at night, at home), topic, affiliation, funding source, device, classroom materials, age group, SDGs, etc. Filters can be combined to construct complex queries. Registered users can also get project recommendations based on

Co-funded by the European Union, Switzerland and the United Kingdom

SciStarter AI recommender, track their contributions across projects and platforms, share their profiles and accomplishments, and leave project reviews.

In addition to project discovery, the platform provides:

- training materials, tools and resources for project owners, public libraries, museums, and schools;
- low-cost tools (e.g., designs for sensors and testing kits) for making observations, recording data, and processing samples;
- project hosting facilities with management of the project participation and project statistics; and
- SciStarter API tools for project owners (SciStarter, 2023b).

The platform also offers a RESTful API to search for and retrieve full projects' metadata in JSON format (query URL structure: `https://scistarter.org/p/finder?format=json&key=API_key&q=)`. The query results are paginated therefore pages need to be requested sequentially until all projects are retrieved.

At present, SciStarter does not support sharing of the projects' data. Projects hosted on the platform can only share aggregated observations on a map.

Figure 4: CitizenScience.gov project finder page.

CitizenScience.gov is an official US government crowdsourcing and citizen science platform for the US federal government and nongovernmental organisations (CitizenScience.gov, 2023). The platform provides a catalogue of federally supported projects and a toolkit of resources to assist in scoping, designing, and managing citizen science projects. The toolkit includes references to external sources rather than resources exclusively developed for CitizenScience.gov. The platform does not host projects and does not support data access. CitizenScience.gov contains 502 live projects, some of which are also listed on other citizen science platforms (e.g., Zooniverse).

Co-funded by the European Union, Switzerland and the United Kingdom

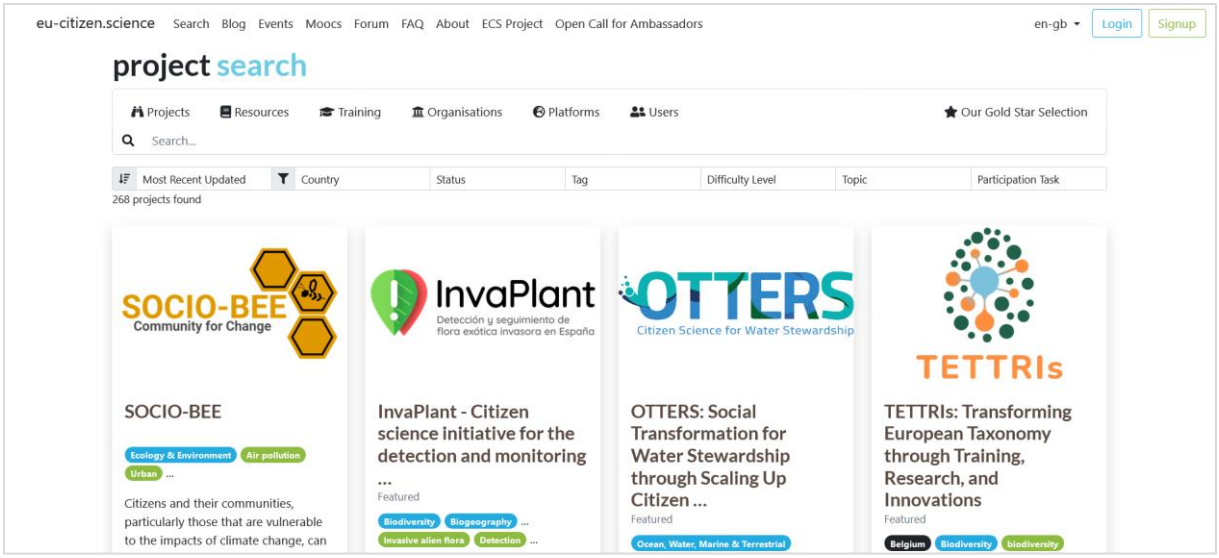


Figure 5: EU- Citizen.Science project discovery page.

EU-Citizen.Science platform was established in 2019, initially funded by the European Union's Horizon 2020 and Horizon Europe Framework Programmes for Research and Innovation (EU-Citizen.Science, 2023a). At present, the platform is supported by a consortium of 14 partners and 9 third parties from across 14 European countries.

The platform is primarily focused on projects within the EU but is not exclusive to Europe. It contains 269 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 in JSON format (EU-Citizen.Science, 2023b).

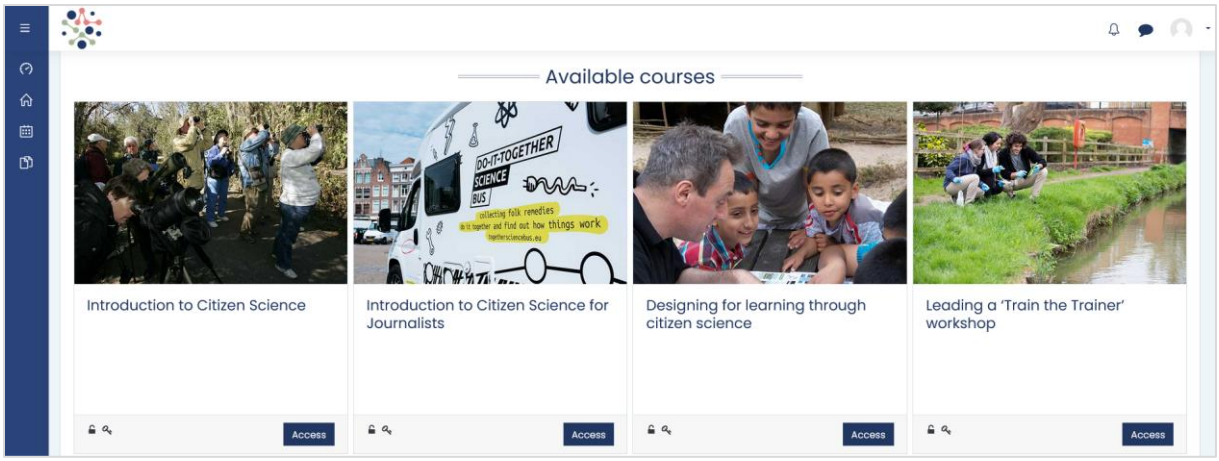


Figure 6: EU-Citizen.Science Moodle training platform.

Since EU-Citizen.Science is most relevant to the EU GDDS, projects' metadata was retrieved using the Swagger API (the API is documented using the Swagger standard, recently rebranded as OpenAPI (EU-Citizen.Science, 2023b)) to analyse the distribution of the projects' status, topics, and participation tasks, and the relationships between topics and participation tasks.

EU-Citizen.Science platform contains 269 project, 219 which are active, 33 are completed, 11 are periodically active, 4 are not yet started, and 2 are on hold (Figure 7). As shown in Figure 8, the most common participation tasks are observation (60 projects), data entry (55 projects), and identification (45 projects) - (NB: a project can include multiple participation tasks). Regarding project topics (Figure 9), Ecology & Environment (102 projects, 13%), Biodiversity (91 projects, 11%), and Education (64 projects, 8%) are the top three of the 29 topics covered. The topics used are of varying scale and can be aggregated, e.g., Birds, Long-

Co-funded by the European Union, Switzerland and the United Kingdom

term species monitoring, Insects & pollinators can all be combined under the Biodiversity theme. As with the participation tasks, a single project can fall under multiple topics. Figure 10 shows a co-occurrence matrix of project topics which can be used to identify topics that are closely related in the EU-Citizen.Science projects. Figure 11 shows a co-occurrence matrix of project topics and participation tasks which can be used to identify the tasks that are most common or those that are never used for a given topic.

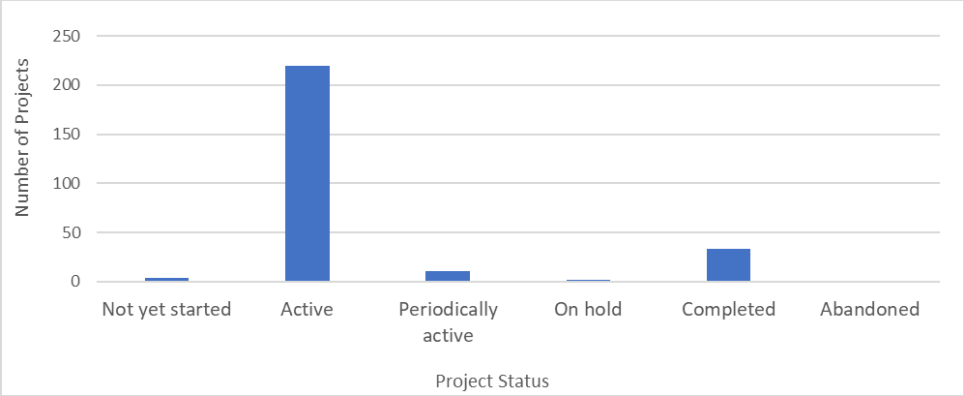


Figure 7: EU-Citizen.Science platform project status.

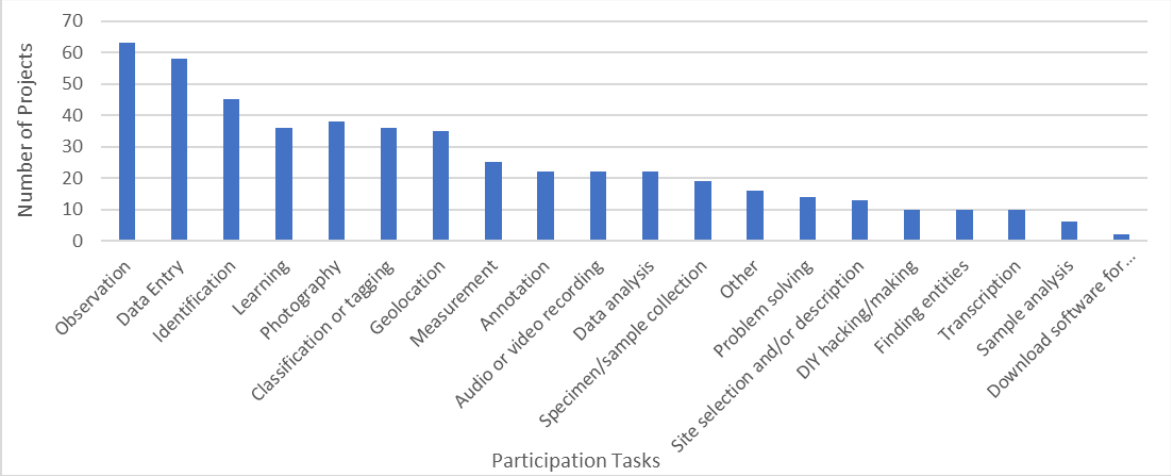


Figure 8: EU-Citizen.Science platform participation tasks.

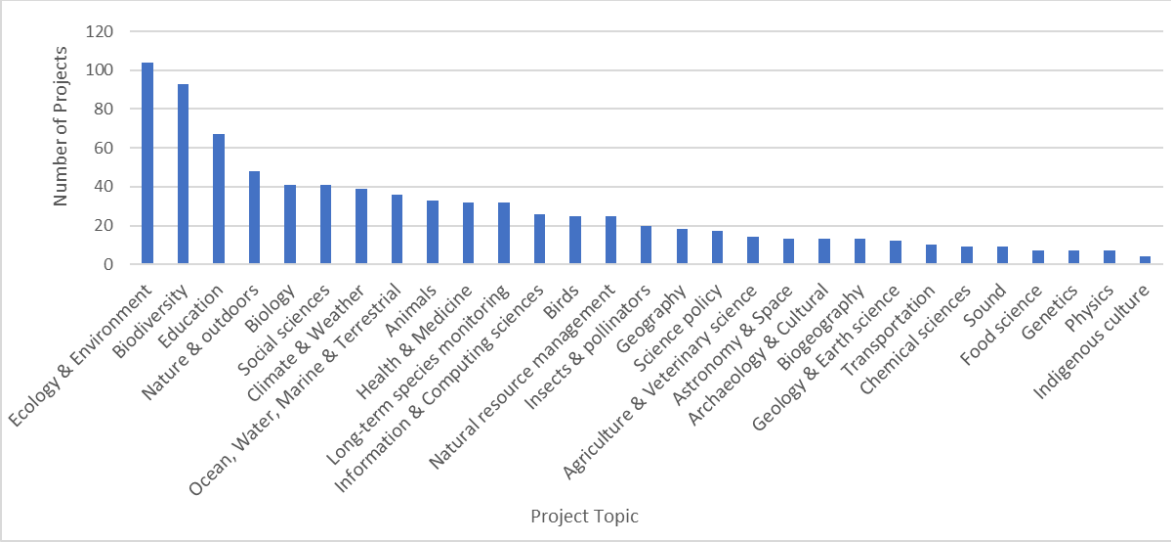


Figure 9: EU-Citizen.Science platform project topics.

Co-funded by the European Union, Switzerland and the United Kingdom

Figure 10: Citizen.Science platform – co-occurrence matrix of project topics.

Co-funded by the European Union, Switzerland and the United Kingdom

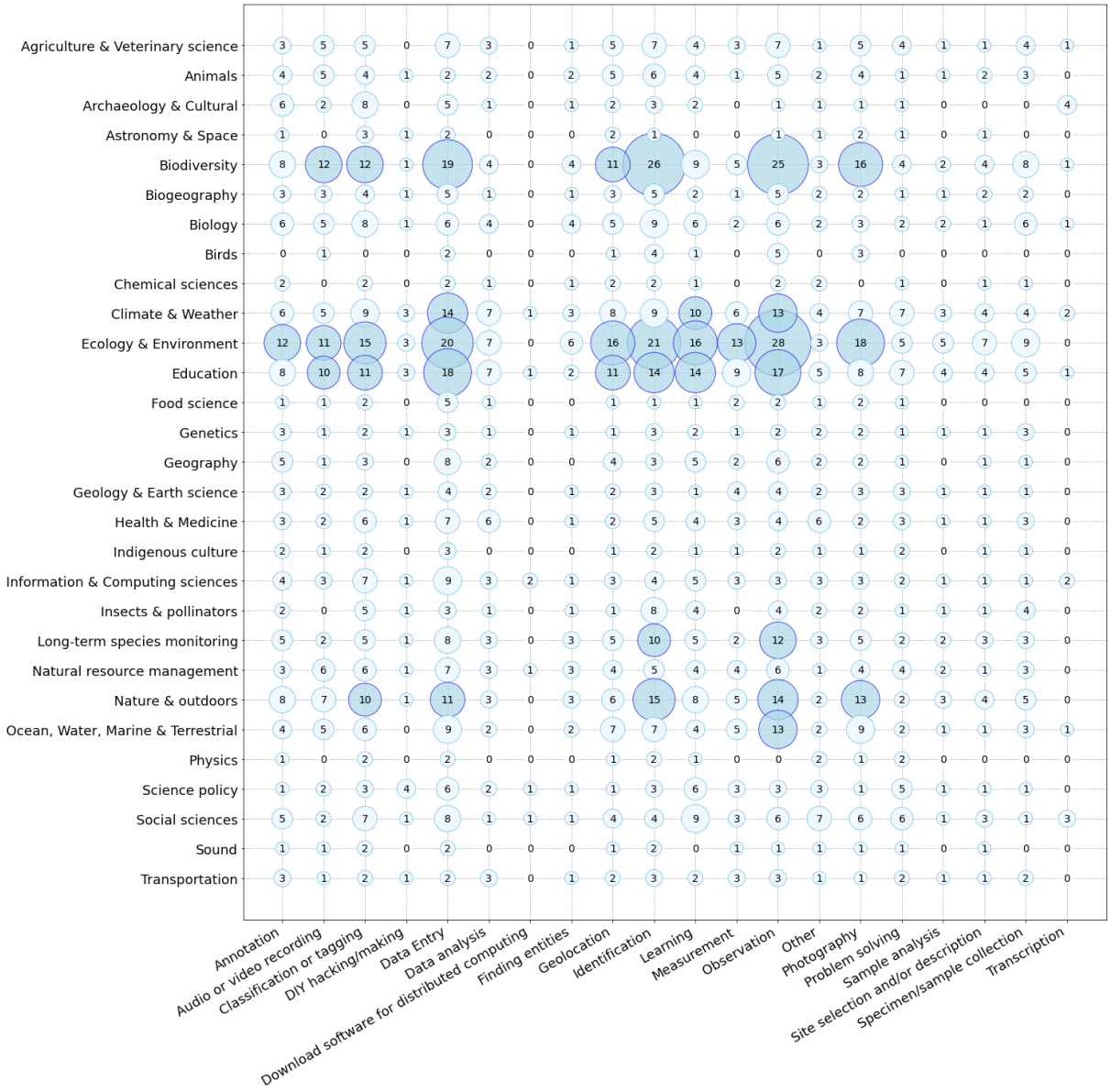


Figure 11: Citizen.Science platform – co-occurrence matrix of project topics and participation tasks.

To identify EU-Citizen.Science platform topics that are potentially relevant to the EU GDDS, topics were mapped to the ***Green Deal themes***. There is no clear definition by the European Commission as what the GD themes are, some EC pages mention a term "*policies*" (EC, 2023b), other talk about "*benefits*" and "*what we are working on*" (EC, 2023c). For the purpose of mapping, the following ***eight EC GD thematic areas*** were used (extracted from EU Commission 2020 press release (EC, 2023d)):

1. Increasing climate ambition
2. Clean, affordable and secure energy
3. Industry for a clean and circular economy
4. Energy and resource efficient buildings
5. Sustainable and smart mobility
6. Farm to fork
7. Biodiversity and ecosystems
8. Zero-pollution, toxic-free environments

Co-funded by the European Union, Switzerland and the United Kingdom

Table 3 presents all topics currently covered by the EU-Citizen.Science platform and their mapping to the GDDS themes and AD4GD project themes. As already noted, the topics are of varying scale. A project can fall under multiple topics.

Table 3: EU-Citizen.Science topics mapped to the GDDS themes and AD4GD project themes.

Topic	No	GDDS Themes	AD4GD Themes
Agriculture & Veterinary science	14	Increasing climate ambition Industry for a clean and circular economy Farm to Fork Biodiversity and ecosystems Zero-pollution, toxic-free environments	Increasing climate ambition Biodiversity and ecosystems Zero-pollution, toxic-free environments
Animals	33	Biodiversity and ecosystems	Biodiversity and ecosystems
Archaeology & Cultural	13	--	
Astronomy & Space	13	--	
Biodiversity	93	Biodiversity and ecosystems	Biodiversity and ecosystems
Biogeography	13	Biodiversity and ecosystems	Biodiversity and ecosystems
Biology	41	Biodiversity and ecosystems	Biodiversity and ecosystems
Birds	25	Biodiversity and ecosystems	Biodiversity and ecosystems
Chemical sciences	9	--	
Climate & Weather	39	Increasing climate ambition	Increasing climate ambition
Ecology & Environment	104	Increasing climate ambition Zero-pollution, toxic-free environments	Increasing climate ambition Zero-pollution, toxic-free environments
Education	67	--	
Food science	7	Farm to Fork	
Genetics	7	--	
Geography	18	--	
Geology & Earth science	12	Industry for a clean and circular economy	
Health & Medicine	32	Zero-pollution, toxic-free environments	Zero-pollution, toxic-free environments
Indigenous culture	4	--	
Information & Computing sciences	26	--	
Insects & pollinators	20	Biodiversity and ecosystems	Biodiversity and ecosystems
Long-term species monitoring	32	Biodiversity and ecosystems	Biodiversity and ecosystems
Ocean, Water, Marine & Terrestrial	25	Biodiversity and ecosystems Increasing climate ambition Clean, affordable and secure energy	Biodiversity and ecosystems Increasing climate ambition

Co-funded by the European Union, Switzerland and the United Kingdom

Nature & outdoors	48	Biodiversity and ecosystems	Biodiversity and ecosystems
Natural resource management	36	Increasing climate ambition Clean, affordable and secure energy Zero-pollution, toxic-free environments	Increasing climate ambition Zero-pollution, toxic-free environments
Physics	7	--	
Science policy	17	-- ?? --	
Social sciences	41	--	
Sound	9	Zero-pollution, toxic-free environments	Zero-pollution, toxic-free environments
Transportation	10	Sustainable and smart mobility Zero-pollution, toxic-free environments	Zero-pollution, toxic-free environments

2.1 EU NATIONAL CITIZEN SCIENCE NETWORKS

In addition to large-scale citizen science project discovery platforms, national citizen science networks are being established and are growing across Europe. The majority of such networks are in the early stage of development, only a few support project discovery, and none yet provide open access to data. Some examples are reviewed below.

The national networks described here are part of the European Citizen Science Association (ECSA) Citizen Science Networks working group (a summary is available in ANNEX II).

Österreich forscht is an Austrian citizen science network founded in 2017 (Österreich forscht, 2023). The network is supported by the University of Natural Resources and Life Sciences, Vienna. Österreich forscht is not domain-specific but has well-defined transparent quality criteria (minimum standards) that the projects must meet before being listed on the project discovery platform. The introduction of Österreich forscht standards criteria sparked active discussions within the international citizen science community on whether this limits the definition of citizen science (Haklay, 2021), yet quality assurance is vital to ensure valid and valuable results. *Quality Criteria for Citizen Science Projects* on Österreich forscht can be found on Zenado at https://zenodo.org/record/1161953.

The project finder platform allows to filter projects by topic, activity, participation, and location. Filters loosely resemble, but do not match the PPSR Core vocabulary. At the time of writing, there are 72 active projects on the platform, although the number of projects is not clearly indicated. The platform only supports project discovery and does not provide access to projects' data.

The OeAD Center for Citizen Science is an Austria-wide citizen science centre coordinated by the Austria's Agency for Education and Internationalisation (Zentrumfuercitizenscience, 2023). The centre offers information about citizen science, funding opportunities, and networking and collaboration.

Virtual Ecosystem for Research Activation (VERA Hub) is a European network that aims to facilitate collaboration and promote citizen science research in Social Sciences and Humanities (VERA, 2023a). The VERA Hub platform is developed by COESO project (COESO, 2023) that has received funding from the European Union's Horizon 2020 Research and Innovation action funding scheme. VERA Hub offers facilities to discover citizen science projects for collaboration (VERA, 2023b). At present, only 14 projects are listed on the platform.

Co-funded by the European Union, Switzerland and the United Kingdom

The Norwegian Network for Citizen Science Forskningsradet is coordinated by Norwegian Research Council and is currently under development (Forskningsradet, 2023). The network will focus on promoting citizen science in Norway, coordinating citizen science activities, and exchanging experience. The network's website is currently available in Norwegian language only.

Citizen Science Netværket is a Denmark-wide citizen science network that focuses on promoting citizen science projects and facilitating networking and collaboration (Citizen Science Netværket, 2021). The network is in very early stages of development.

Citizen Science Italy is in its very early stages of development and is awaiting potential funding from the National Biodiversity Future Center (Citizen Science Italy, 2023). The network will focus on coordinating citizen science initiatives in Italy and aim to achieve formal recognition of citizen science activities at the national level.

Rede Portuguesa de Ciência Cidadã / CC.pt is a Portuguese citizen science network that is run by volunteers and does not yet have sources of funding (CienciaCidada, 2023). The website is only available in Portuguese and currently lists only 3 external citizen science projects.

Iedereen Wetenschapper is a citizen science network in Flanders and the Netherlands (IedereenWetenschapper, 2023). The network provides a citizen science project discovery platform for all disciplines but is only available in Dutch language. The project discovery page only displays the project list once filtering criteria is applied; it does not provide information on the total number of projects registered on the platform. The projects can be searched by keywords, task duration, topic, location, activity, intended participant type, and past/future projects. Iedereen Wetenschapper works in close collaboration with Citizen Science Vlaanderen.

Scivil (Citizen Science Vlaanderen) is a citizen science network that works towards promoting citizen science and connecting stakeholders in Flanders and Belgium (Scivil, 2023). The network provides a citizen science project discovery platform that supports both Dutch and English languages. The platform does not provide clear information on the number of projects listed; it supports filtering by topic and year, although filtering by year causes website error.

Citizen Science Nederland (CS-NL) is the Netherlands citizen science network that was launched in 2022 by the National Program Open Science (CS-NL, 2023). The network is currently in early development and is awaiting funding from the Open Science Regieorgaan in the Netherlands.

Citizen Science Center for Nature, Sustainability and Digitalization is an 18-month project funded by Deutsche Bundesstiftung Umwelt and coordinated by Musem of Natural History Berlin (Museum Fuer Naturkunde, 2023). The project aims to establish a citizen science centre in Germany that will support nature conservation and sustainability citizen science projects to overcome digitalisation challenges and increase digital and technological competences. The project highlights that recent developments in technologies and artificial intelligence can be leveraged by citizen science projects.

Medborgar Forskning.se is a Swedish citizen science network that offers tool and guidelines to citizen science researchers and other stakeholders (Medborgarforskning, 2023). The networks' website was launched in 2021 by the ARenas for Cooperation through citizen Science (ARCS) project that received funding from Vinnova (Sweden's innovation agency), the University of Gothenburg, the Swedish University of Agricultural Sciences, and Umeå University. Medborgar Forskning.se uses EU-Citizen.Science platform as a portal for citizen science projects discovery in Sweden.

CitizenScience.cz is a Czech Republic citizen science network that is coordinated by the Czech Academy of Sciences and Tomas Bata University in Zlin (Ústav Geoniky, 2021). The network offers a discovery platform and a database for citizen science projects in Czech Republic but is only available in Czech language. The platform does not clearly indicate the number of listed projects and only supports search by keywords.

Co-funded by the European Union, Switzerland and the United Kingdom

The majority of the national citizen science networks described above are still in active development. Six of these networks provide their own project discovery platforms and one supports project discovery through EU-Citizen.Science. At present, however, none of the platforms provide access or links to projects' data.

3 DATA AGGREGATION PLATFORMS

Data aggregation platforms are platforms that collect domain-specific data and facilitate data sharing. In this section, we provide examples of data aggregation platforms that could potentially serve as data sources for the EU GDDS. We also discuss other examples of citizen science aggregation platforms for completeness. It is worth mentioning that these data aggregation platforms are replicating the data. The alternative could be data federation or data discovery platforms, which has been detected as a gap in the ecosystem of currently-available tools. Cos4Env (Cos4Env, 2022) and Cos4Bio (Cos4Bio, 2022) were two attempts at data federation developed in the context of the COS4CLOUD project. Another initiative that aims to achieve a federation of CitSic projects is currently underway in the CitiObs project for Air Quality measurements (CitiObs, 2023.).

Citizen science projects or platforms can follow either a bottom-up or top-down approach (Liu *et al.*, 2021). In a ***bottom-up approach***, standards are loosely defined, and all members can equally participate in decision making. The main benefit is flexibility and natural shaping of standards from diverse community contributions. An example is ***OpenStreetMap*** (described in more detail below) where "standardised" map object tags (or descriptions) emerged from free-text community contributions. Flexibility can also be a disadvantage, since communal harmonisation and decision-making are slow, and a non-standardised approach affects interoperability when it is impossible to know what to expect from data. In addition, funding and platform stability can be challenging to maintain. In a ***top-down approach***, a governing body or a funder make executive decisions and standards are rigorously defined. The benefits of this approach are standardised protocols and data formats that support interoperability – users and machines know what to expect. An example is the **GBIF platform** for biodiversity data. One drawback of this approach is the fact that valuable knowledge can be lost if concepts are not represented in a strictly-defined data model.

3.1 DATA AGGREGATION PLATFORMS RELEVANT TO THE GDDS

In this section, we discuss examples of citizen science initiatives, namely GBIF, OpenStreetMap and Sensor.Community, each of which closely aligns with at least some of the FAIR data principles.

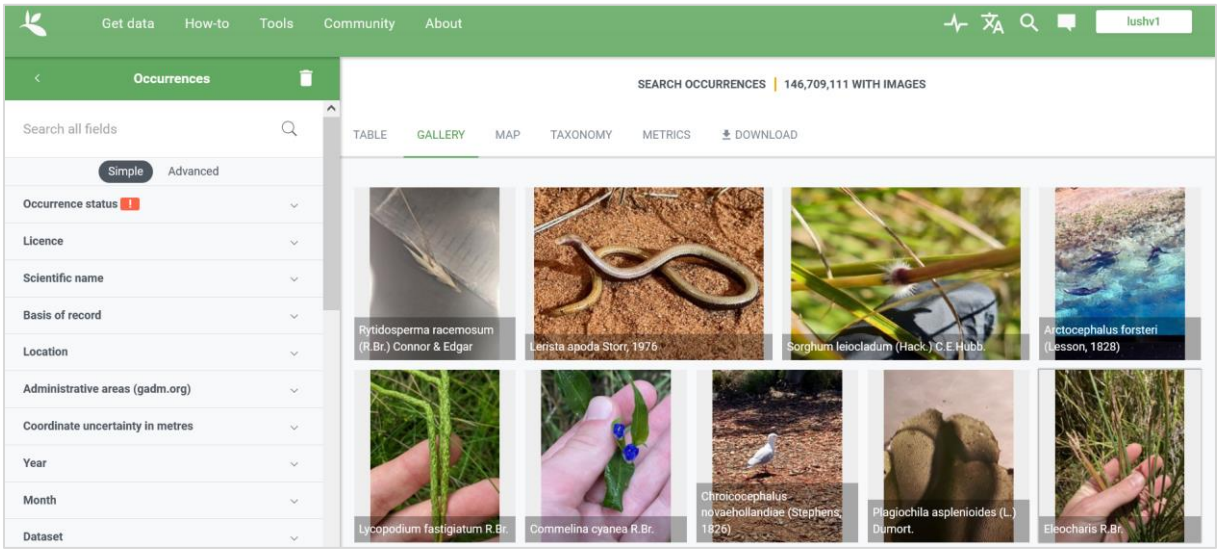


Figure 12: The Global Biodiversity Information Facility (GBIF) home page.

The Global Biodiversity Information Facility (GBIF) is an international network that promotes and facilitates free and open access to biodiversity data from across the globe (GBIF, 2023a). GBIF was established

in 2001 through a Memorandum of Understanding between participating governments and is now funded by agencies from national governments that have voting rights. GBIF accepts data from diverse sources including citizen science initiatives (e.g., iNaturalist and eBird projects that are described later in this report).

GBIF provides facilities to search for species occurrences data, species in the GBIF taxonomy data, and biodiversity datasets. Species occurrences data can be searched using the following simple filters (advance filtering is also available): keywords, occurrence status, licence, scientific name, basis of record, location, administrative areas (gadm.org), coordinate uncertainty in metres, year, month, dataset, country or area, continent, issues and flags, media type, publisher, institution code, collection code, catalogue number, type status, IUCN Global Red List Category). Species taxonomy filters include keywords, rank, status, higher taxon, and issues and flags. To search for biodiversity datasets, GBIF provides the following filtering functionality: keywords, publisher, host, publishing country or area, project, and licence. GBIF is partner in the B3 sister project where the project will contribute dynamic data cubes based on the platform's current data (B-Cubed, 2023).

The platform contains 2,364,723,400 species occurrence records and 86,951 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 is accompanied by a DOI that, in accordance to the licence, must be cited when using the data; this increases transparency and reproducibility by recording the provenance of the data.

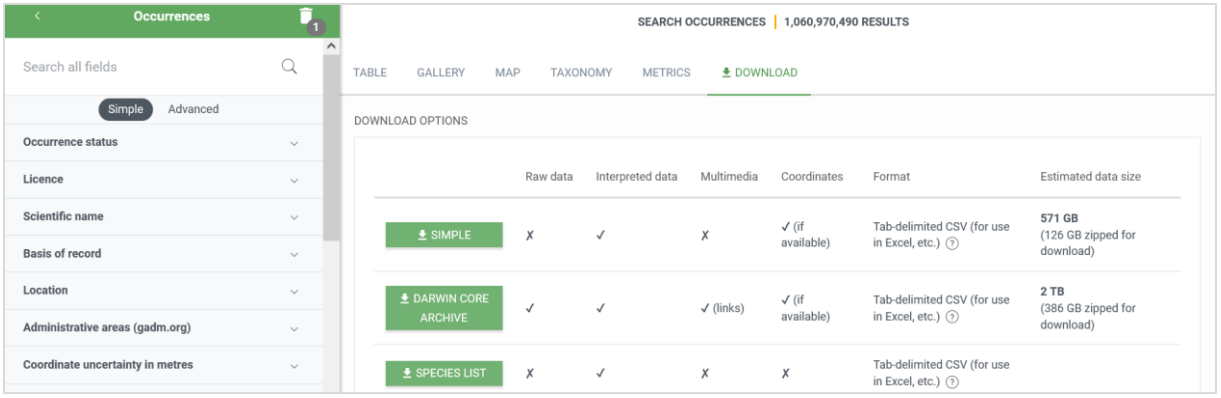


Figure 13: GBIF data download page.

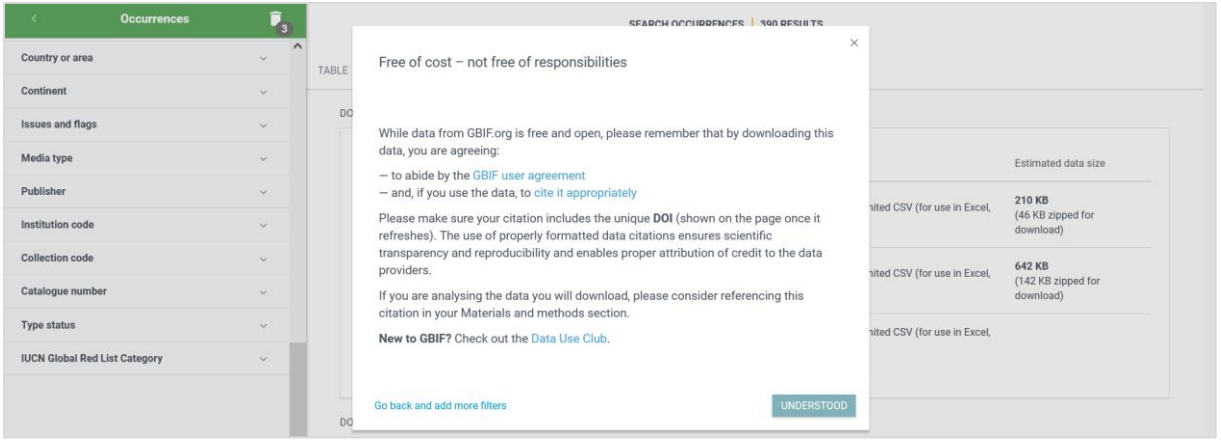


Figure 14: GBIF data download page – licence statement.

Co-funded by the European Union, Switzerland and the United Kingdom

gbifID	datasetKey	occurrenceID	kingdom	phylum	class	order	family	genus	species	infraspecificEpithet	taxonRank
978332733	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD:OBS233006732	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES
977962157	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD:OBS231755412	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES
976917267	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD_NZ:OBS224521081	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES
964739786	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD:OBS210057946	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES
943414365	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD:OBS186133229	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES
935171905	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD:OBS176044459	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES
821232739	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD:OBS150671067	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES
819761783	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD_CAN:OBS155289873	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES
818297927	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	URN:catalog:CLO:EBIRD_NZ:OBS173165647	Animalia	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	Apteryx rowi		SPECIES

Figure 15: Example GBIF data file.

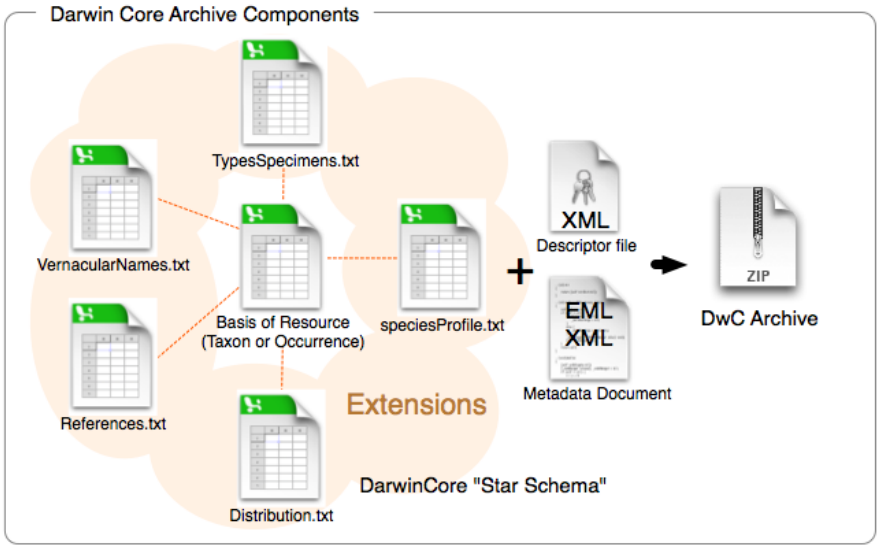


Figure 16: Darwin Core Archive components (Gcube, 2013).

While Darwin Core is the required format for GBIF published data, there is a 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 (GBIF, 2023b) and is actively supporting the initiative to evolve their biodiversity data model (GBIF, 2023c).

In addition to GBIF central data discovery and access platform, GBIF offers facilities to host external portals in a form of fully customisable websites to support national, institutional, regional and thematic initiatives.

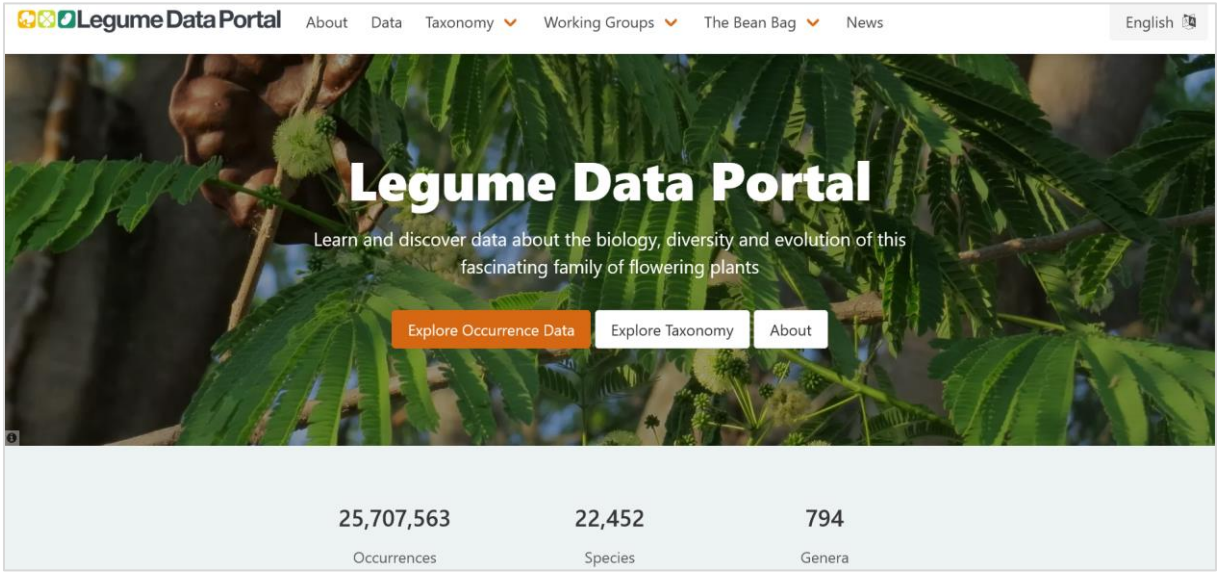


Figure 17: Example GBIF hosted portal.

Co-funded by the European Union, Switzerland and the United Kingdom

GBIF offers a set of data formatting, documentation, and validation tools to help data providers to publish and share biodiversity data via the GBIF portal.

The Darwin Core Archive Assistant facilitates publishing of biodiversity data using the Darwin Core Archive format (DwC-A) (required format for publishing data to GBIF) (GBIF, 2023d). The tool helps to produce a single self-contained dataset for species occurrence or taxonomic (species) data. While facilitating structured standardised data representation, the Darwin Core Archive Assistant does not offer automated data formatting and can be laborious to reformat the existing data.

Figure 18: GBIF - the Darwin Core Archive Assistant.

The Integrated Publishing Toolkit (IPT) facilitates publishing and sharing of biodiversity datasets through the GBIF network (GBIF, 2023e). It supports publishing of various types of data resources in Darwin Core format and metadata for these datasets in EML format. Only open data is supported; sensitive data needs to be removed prior to publication.

The GBIF Validator Tool allows to validate biodiversity datasets before uploading these to GBIF (GBIF, 2023f). The Validator Tool accepts zip-compressed Darwin Core Archives, IPT Excel templates, and simple CSV files containing Darwin Core and returns a report on the syntactical correctness and the validity of the dataset content.

While Darwin Core is the required format for GBIF published data, there is a 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 (GBIF, 2023b) and is actively supporting the initiative to evolve their biodiversity data model (GBIF, 2023c).

The success of GBIF in becoming the largest Open biodiversity data provider lies not only in developing a stable software platform, but in providing standardised but evolving data and metadata standards, best practices documents, and technical tools for structuring and uploading biodiversity data. '*Current Best Practices for Generalizing Sensitive Species Occurrence Data*' document (Chapman, 2020), for example, helps data providers to publish as much Open data as possible without endangering sensitive species, since only open data is accepted by GBIF. These resources make the platform more accessible for a wide range of stakeholders and ensure data Openness, correctness, and interoperability.

Co-funded by the European Union, Switzerland and the United Kingdom

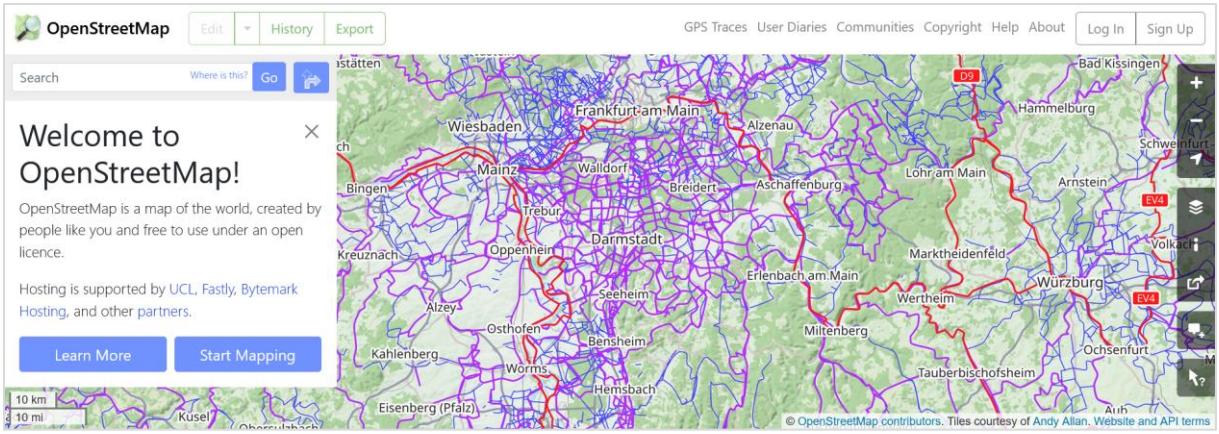


Figure 19: OpenStreetMap platform page.

OpenStreetMap (OSM) is a collaborative platform and project that aims to create an editable, open access map of the world from contributions by citizens (OSM, 2023a). A community of volunteer contributors 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. OSM was launched in 2004 as part of a MSc project dissertation by Steve Coast and has since grown into the most extensive and detailed open source mapping platforms available. The platform is financed by regular donations, intermittent fundraising appeals and the OpenStreetMap Foundation memberships. It is hosted with the support from University College London, Fastly, Bytemark Hosting, 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 key features and review the most prominent services and tools that utilise OSM data.

The key features of OSM include:

Spatial database: The core purpose of OSM is to map the physical features of the world (e.g., roads, buildings, addresses, land use) and provide up-to-date free, open access geospatial database.

Collaborative data collection: OSM has a vast community of volunteers who work together to gather and maintain the geospatial data. Anyone can become a contributor and add details about their local area or areas of interest. OSM relies on local knowledge and local experts to verify the quality of the data.

Open data: All the data in OSM is open and freely available under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF) (OSM, 2023b). This means that the map data can be accessed, used, and shared by anyone for various purposes without restrictions.

Continual maintenance: OSM does not rely on periodic updates and is continually maintained by the community to reflect changes. Contributors update the status of existing features and create new points of interest to keep the map current and accurate. Updates to every OSM element are stored so changes can be traced back to the creation of the object. Snapshots of data can be extracted to assess the changes to areas or features of interest.

Global coverage: OSM aims to provide comprehensive coverage data for the entire world in a unified tagging schema. While some areas may have more detailed mapping due to the presence of active contributors, the goal is to provide at least basic coverage for every region.

Custom mapping and applications: The open nature of OSM data allows developers and organisations to create custom maps and develop location-based applications using the OSM data. The only requirement is that the results are distributed under the same ODbL license as the OSM data used.

Supporting humanitarian and crisis response efforts: OSM has been instrumental in supporting disaster response and humanitarian efforts by providing up-to-date maps of affected areas where commercial

mapping services might not have sufficient data. Historical data plays an important role in crisis response where changes to affected areas can be extracted and evaluated.

A vast and ever-evolving range of third-party applications, tools and services are developed using OSM data (OSM, 2023c), many of which are relevant to facets of the Green Deal (see also section 7.1 for examples of GDDS-relevant research applications). Commercial companies use OSM data for mapping services (***Geofabrik***), navigation (***Mapbox***, ***Mapzen, OSMAnd***), live traffic updates and road conditions (***MapQuest***), street-level images (***Mapillary***), data visualisation and cartography (***Stamen***), geospatial analytics (***CampToCamp***), etc. 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 the high value of VGI.

The core function of OSM is to collect, maintain, and distribute ***an open global geospatial database***, rather than produce cartographic products and maps (Mooney and Minghini, 2017). The OSM conceptual data model of the physical world consists of three basic elements: ***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 (OSM, 2023d). 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 essential semantic meaning to each element in the database. There are a lot of resources to guide users in identifying appropriate tags and understanding tag usage, e.g., ***TagInfo*** site offers statistics on the entire OSM tags database (TagInfo, 2023). OSM contains over 24 billion GPS points, over 8.5 billion nodes, close to 1 billion ways, and over 11 million relations (Planet OSM, 2023a).

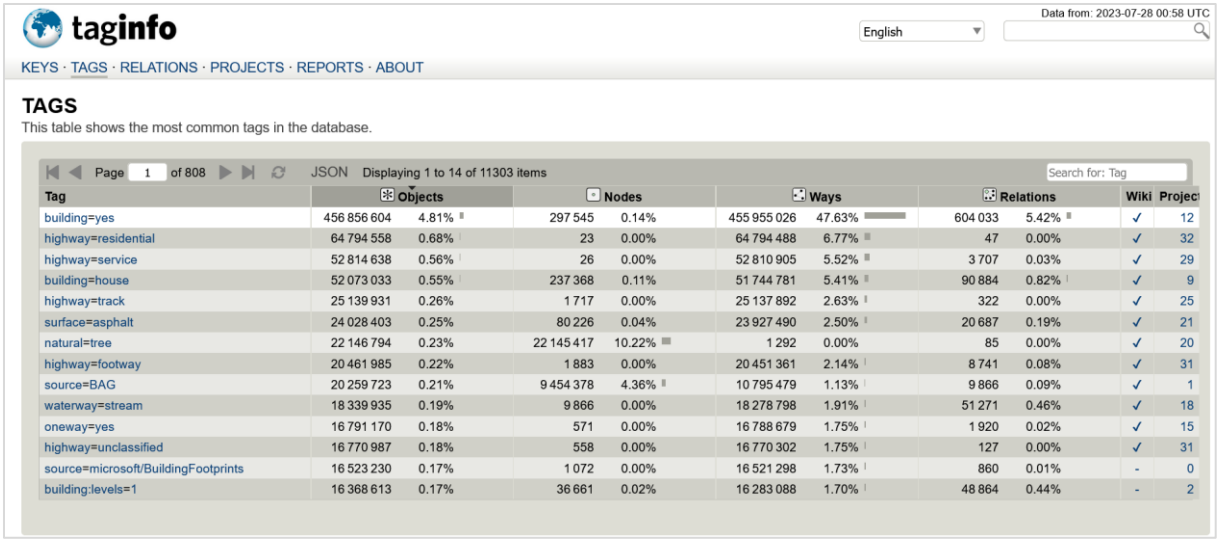


Figure 20: TagInfo website for OpenStreetMap tags statistics.

There are a number of ways in which OSM data can be downloaded (OSM, 2022a). Planet OSM (Planet OSM, 2023b) offers a complete copy of OSM database that is updated weekly and is downloadable in a compressed XML file and an OSM native Protocol Binary Format (PBF). Planet OSM additionally offers download of full OSM editing history in XML and PBF formats (Planet OSM, 2023c). Both datasets are huge (full planet dataset is over 120GB and history dataset approaching 200GB) and can be slow to download. For faster download that does not overload the primary OSM server, ***OpenMapTiles*** project offers full planet dataset served from multiple mirrors at once (OpenMapTiles, 2023).

Co-funded by the European Union, Switzerland and the United Kingdom

Regional datasets can be obtained for free from ***Geofabrik*** download server (Geofabrik, 2023). User can select a continent and a county of interest and download data in PBF and ESRI compatible shape file formats. Geofabrik server data is updated daily but omits personal information such as user names, user IDs and changeset IDs of the OSM objects due to data protection regulations.

OSM also provides unfiltered access to its raw data via ***Overpass API*** (OSM, 2023e). Overpass API enables custom querying of data with specific tags, location and temporal resolution. API's main strength is filtering flexibility, but data is returned in XML format that is not readily compatible with most GIS tools. Overpass Turbo can be used to obtain data in GeoJSON format (OSM, 2020).

Small data extracts can be exported in .osm format directly from the OSM platform by navigating to ***Export*** tab and selecting the desired area.

OSM RESTful ***Editing API*** (OSM, 2023f) can also be used to retrieve raw data from the OSM database programmatically. The API supports developers and applications in creating, reading, updating and deleting OSM data programmatically, however large queries are not allowed.

Some land cover products have been derived by different teams from the OSM such as Heidelberg University (UHEI) (HeiGit, 2023) and the University of Coïmbra (University of Coïmbra, 2023). These derived products could also be of interest for the GDDS.

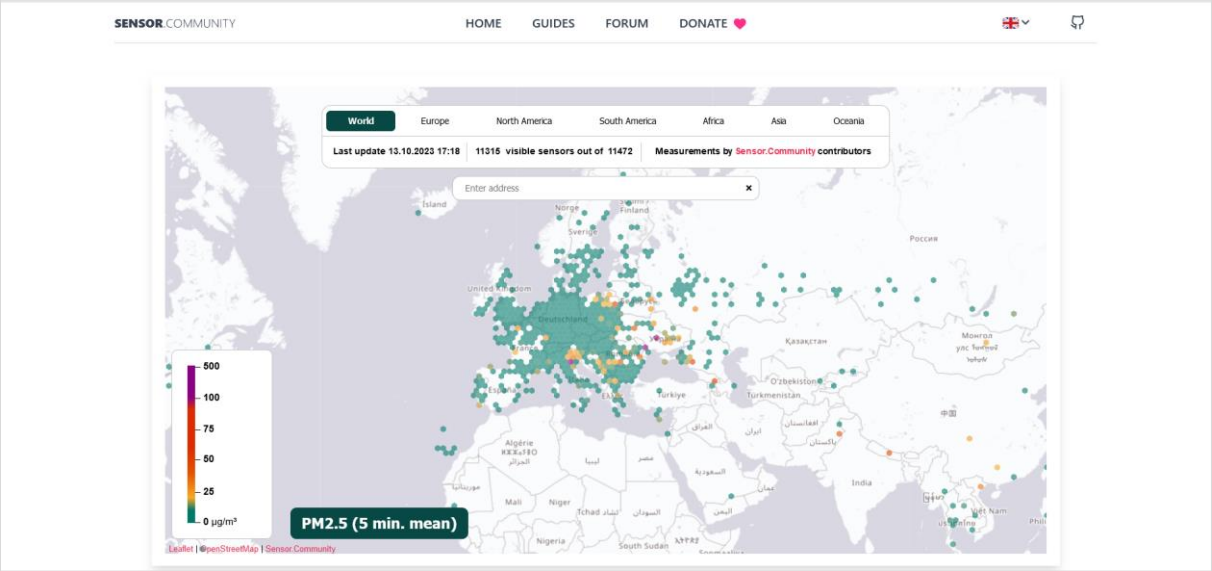


Figure 21: Sensor.Community home page.

Sensor.Comunity, 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 (Sensor.Comunity 2023a). 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 most sensors are concentrated in Europe). The project is currently supported by volunteers and voluntary donations.

The Sensor.Community's goal is to raise awareness about air pollution and its potential health and environmental impacts and enable citizens to actively participate in monitoring and improving air quality in their communities. The project also aims to create a comprehensive dataset that can be used for research, advocacy, and policymaking related to air quality improvement. Some examples of applications of Sensor.Community data include a ***Samen voor Zuivere Lucht*** platform that combines Sensor.Community data with official data sources (Samen voor Zuivere Lucht 2023), ***Samen Meten*** (Measure Together) portal that harvests Dutch data from Sensor.Community database (SamenMeten 2023), and ***HackAir*** platform that

Co-funded by the European Union, Switzerland and the United Kingdom

uses Sensor.Community data to generate information on air quality, thermal comfort, and the probability of forest fires in Europe (HackAIR, 2023).

The sensor kits can be assembled to measure environmental factors (temperature, pressure, humidity), particulate matter pollutants (PM10 and PM2.5), and/or noise, and once configured, can be registered with the platform. The aggregated results are displayed on a live map from 12,952 active sensors in 78 countries with over 23 billion data points. The map offers the following data visualisation layers: NO2 (month. mean), sensors at reference stations, temperature, relative humidity, pressure, noise, AQI US, PM10 (EU, 1 h. mean), PM2.5 (EU, 1 h. mean), PM10 (WHO, 1 h. mean), PM2.5 (WHO, 1 h. mean), PM10 (5 min. mean).

Historic data from 2015 onwards can be downloaded from the Sensor.Community Archive automatically by writing custom scripts (Sensor.Community 2023b). 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, but controlled vocabularies such as WHO Glossary on Air Pollution (WHO, 1980), DEFRA Air Pollution Glossary (DEFRA, 2023), Eionet Data Dictionary (Eionet, 2023), or others could be used to semantically enrich Sensor.Community data and improve its semantic Interoperability.

GBIF, OSM, and Sensor.Community are strong candidates for inclusion into the GDDS. However, these are only three examples of large-scale initiatives that can offer FAIR citizen science data. There are thousands of ongoing and completed citizen science projects that might not align with these well-established initiatives but might hold data of potential value for GDDS analyses, visualisations and decision-making. While the challenges of discovering and using these data persist, the provision of software tools, semantic resources, metadata, and data standards that can support citizen science projects in collecting, structuring, documenting, and sharing data will contribute towards making citizen science data FAIR.

3.2 OTHER DATA AGGREGATION PLATFORMS

In this section, we present examples of aggregation platforms that are not directly relevant to the EU GDDS but offer valuable insights into how citizen science data is collected and shared.

Three of the platforms discussed are linked to GBIF. ***The Atlas of Living Australia (ALA)*** works in collaboration with GBIF and shares the GBIF codebase. It provides Australian biodiversity data to GBIF via the BioCollect data collection tool. ***iNaturalist*** and ***eBird*** citizen science platforms also supply datasets to GBIF.

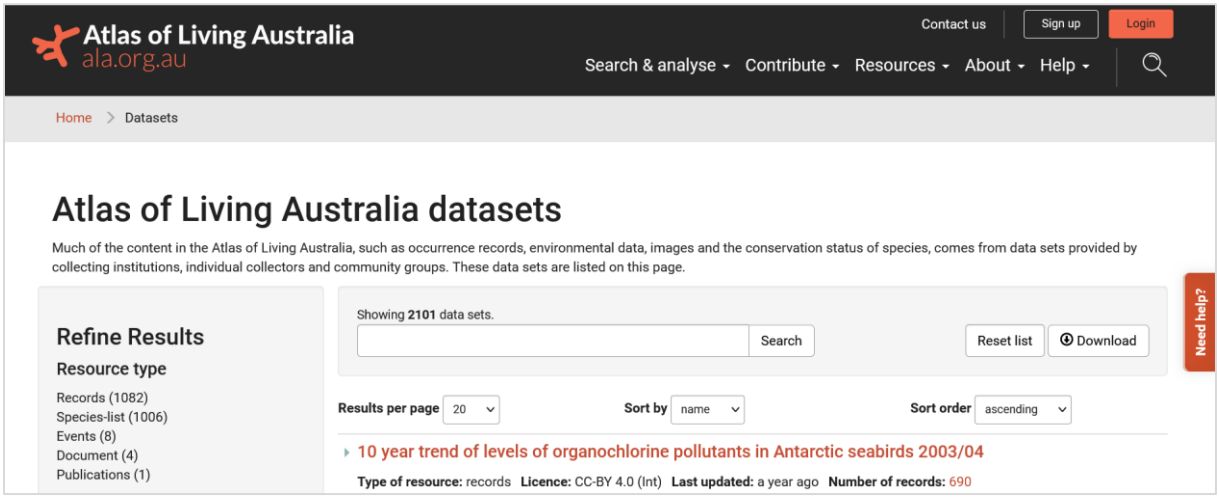


Figure 22: Atlas of Living Australia (ALA) datasets search page.

Co-funded by the European Union, Switzerland and the United Kingdom

The Atlas of Living Australia (ALA) is an open infrastructure that pulls together and provides access to Australia's open data on biodiversity (ALA, 2023). The ALA was launched in 2010 and is now supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) and is hosted by the Commonwealth Scientific and Industrial Research Organisation (CSIRO). It is not specific to citizen science projects and data but supports data collection and sharing by citizen scientists.

The ALA is a database of open biodiversity data (data is hosted on the platform) and provides a comprehensive set of tools and filtering functionality to enable data discovery. Search and filtering can be done by keywords, resource type, license type, integration status, content type, and institution.

The ALA uses the Darwin Core set of standards for their biodiversity data. At the time of writing, the ALA contains 124,631,965 records and 2,101 datasets.

In addition to facilitating open data access, the ALA offers an advance data collection platform for biodiversity science – BioCollect – which is designed for scientists, ecologists, natural resource managers, and citizen scientists. The platform offers a generic BioCollect mobile app for citizen science projects, provides iOS and Android mobile app code via GitHub as a baseline for custom mobile apps, and harvests weekly data submitted by iNaturalist Australia mobile app. Since September 2019, BioCollect data is also accessible via GBIF (GBIF, 2023g).

While the ALA is not citizen science exclusive and only collects and hosts data for Australia, it is a good example of an end-to-end solution for data collection, processing, storage, visualisation, and sharing.

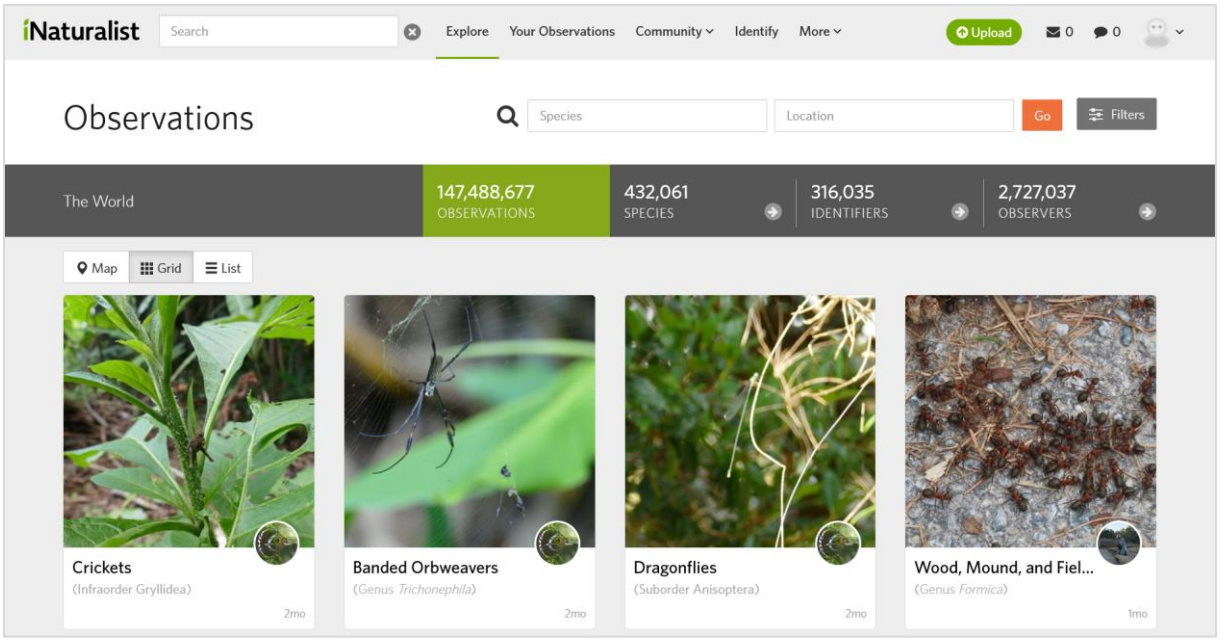


Figure 23: iNaturalist platform – observations page.

iNaturalist is an online platform and a community dedicated to the documentation and sharing of observations of biodiversity (iNaturalist, 2023a). It was established in 2008 by Ken-ichi Ueda and is now a joint initiative by the California Academy of Sciences and the National Geographic Society. iNaturalist serves as a social network and crowdsourcing platform for nature enthusiasts, citizens, scientists, and anyone interested in the natural world.

The platform allows registered users to upload photos and observations of plants, animals, and other organisms. These observations can be geotagged and annotated with relevant information such as species identification, behaviour, habitat. Other users can then help improve or validate the observations by

Co-funded by the European Union, Switzerland and the United Kingdom

identifying the species or providing additional information. This creates a collaborative and educational environment.

iNaturalist offers a mobile app for Android and iOS devices, making it convenient for users to document and share their observations on the go. The app allows for offline data collection and synchronisation when internet connection becomes available. Almost all iNaturalist software is open source and is available on GitHub (iNaturalist, 2023b).

The platform employs image recognition technology and artificial intelligence algorithms to suggest identifications for the uploaded observations. It also provides access to extensive taxonomic databases and field guides to assist users in identifying species accurately.

Although all observations contributed to the platform are public data, iNaturalist supports geoprivacy and taxon geoprivacy. Individual observers can opt in for geoprivacy to obscure their location for personal reasons. For taxonomic groups that might be threatened by disclosing their location, iNaturalist automatically applies taxon geoprivacy. To make observations more valuable for local institutions and authorities, individuals can join their corresponding localised iNaturalist Networks (iNaturalist, 2023c). On joining a localised network, the individual's email address and the true locations of the observations (that are automatically hidden or obscured from public) become available to local institutions that work towards protecting local biodiversity. Pictures are protected by a licence model that is selected by each user.

The platform allows download of observations data by constructing an observations query using supported filters (Figure 24). Observation URLs copied from the website can also be used. Up to 200,000 observations can be downloaded per query. Observation data exports are returned in CSV format; URLs to media files such as images and sound are included in the CSV file.

For bulk data exports, it is encouraged to use alternative download options, in order not to affect iNaturalist platform performance. iNaturalist Research Grade licensed observations can be downloaded via a ***GBIF*** Occurrences Search (select filters: Dataset --> iNaturalist Research-grade Observations); also available as GBIF dataset (GBIF, 2023h). These exports are accompanied with a DOI to support traceability and citation. Another option is to use ***DarwinCore Archive (DwC-A) for GBIF*** (GBIF, 2023i). This is a very large zip file containing records in CSV format that is generated by iNaturalist for GBIF to ingest and is updated weekly.

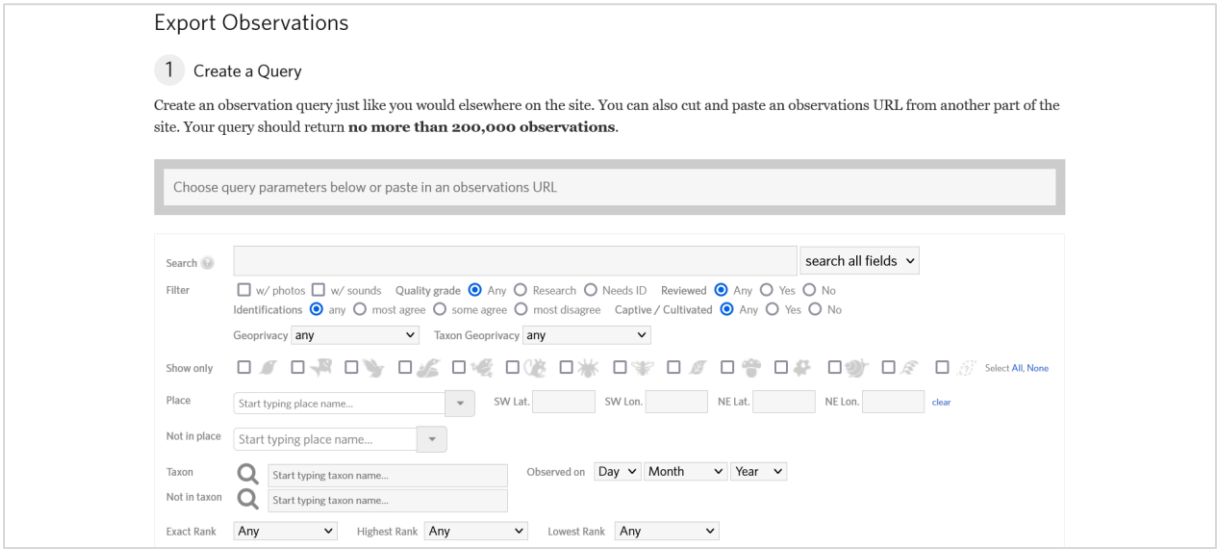


Figure 24: iNaturalist data export page.

Co-funded by the European Union, Switzerland and the United Kingdom

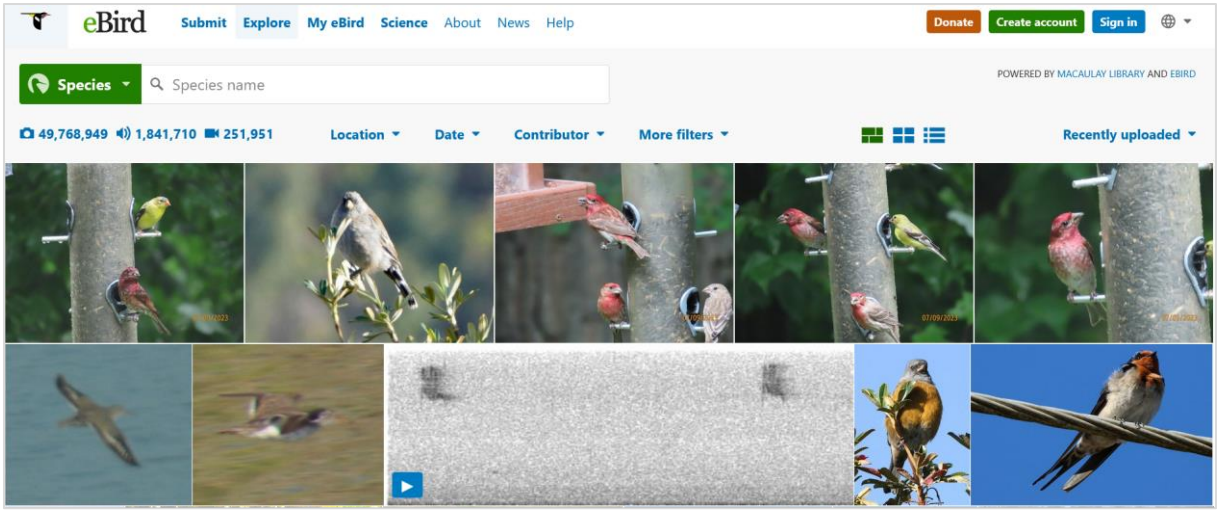


Figure 25: eBird platform home page.

eBird is an online platform and a citizen science project that focuses on bird observation data collection (eBird, 2023). It was launched in 2002 by the Cornell Lab of Ornithology and the National Audubon Society and has since become one of the largest and most comprehensive bird observation resources.

The main purpose of eBird is to encourage birdwatchers and bird enthusiasts to contribute their observations to a central database. Registered users can contribute their observations using the eBird website or use the mobile app to record and submit their bird sightings. Data collected includes information such as the species observed, location, date, and additional details like behaviour, abundance, and breeding activity.

The eBird platform provides the following key functions:

Data collection: eBird offers most comprehensive bird observation database invaluable for scientific research, conservation planning, and monitoring changes in bird populations over time.

Citizen science: eBird is a citizen science project that encourages and facilitates public participation in scientific research.

Personal tracking and tools: The platform facilitates registered birdwatchers to keep track of their personal birding lists, set goals, maintain their own records, explore and visualise their data, create custom checklists, generate personal statistics and reports, and download their data.

Birding hotspots and rarity alerts: eBird provides insights into popular locations for birdwatching and allows birdwatchers to discover new birding spots, learn about recent sightings, and find information about rare or unusual bird species in specific areas.

Conservation and research: Data collected via eBird contributes to bird conservation efforts and research projects and helps to identify important bird areas, monitor migratory patterns, study population trends, and inform conservation strategies.

eBird offers access to its open data as file downloads and through APIs in various formats including raw observation data and processed datasets. Available downloads include Basic Dataset, Status and Trends, Observational Dataset, and user's personal eBird data. To download open data, interested party needs to submit a request with a short description of how eBird data is intended to be used.

eBird data can also be downloaded via a ***GBIF*** Occurrences Search by selecting Dataset --> EOB eBird Observation Dataset filters; also available as GBIF dataset (GBIF, 2023j). While these exports are accompanied with a DOI to support traceability and citation, eBird iteratively corrects data submitted to GBIF (deletes and resubmits datasets). This improves data quality, but creates an issue with DOI consistency and reproducibility of data analysis.

Co-funded by the European Union, Switzerland and the United Kingdom

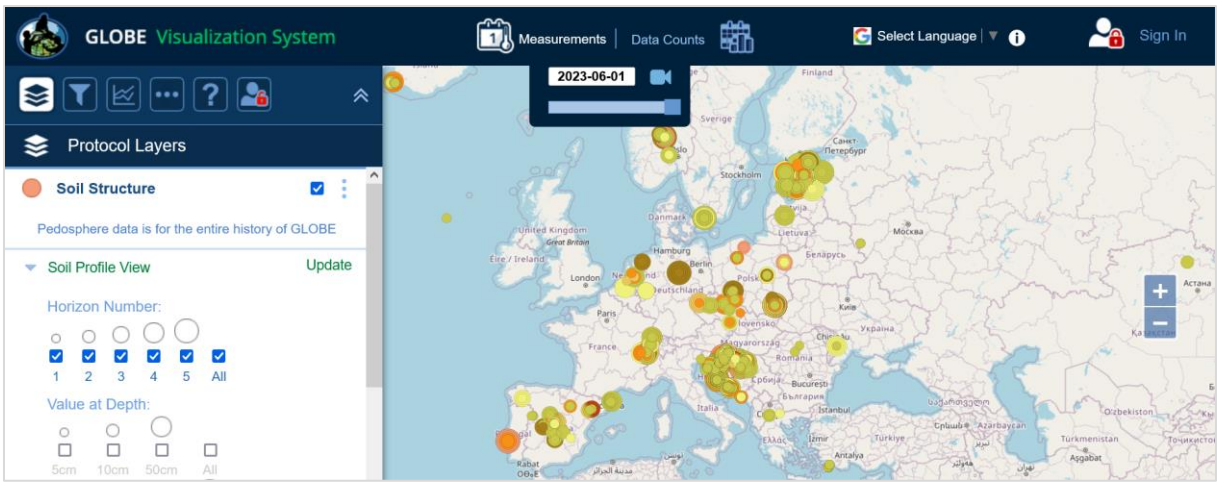


Figure 26: The Global Learning and Observations to Benefit the Environment (GLOBE) platform.

The Global Learning and Observations to Benefit the Environment (GLOBE) program is an international science and education program that provides a framework and resources for students, teachers, scientists, and citizens to make observations, collect data, and contribute to scientific research to better understand the Earth system and global environment (GLOBE, 2023a). The program was launched in 2015 and is currently supported by various U.S. governmental agencies, including the U.S. National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration, the U.S. Department of State, and the National Science Foundation.

The GLOBE Program involves hands-on learning and field investigations, allowing students and citizen scientists to develop their scientific skills and understanding of Earth's systems. It promotes cross-disciplinary learning, combining elements of science, technology, engineering, and mathematics (STEM) education with cultural and language studies.

The program provides a framework and resources for students and participants to make observations, collect data, and contribute to scientific research in various environmental fields. Participants have access to training materials, protocols, and tools for data collection, measurement, and analysis. These resources help ensure data quality and consistency across different locations and projects. The program also fosters international collaboration, encouraging participants to engage in global partnerships and share their findings with other GLOBE schools and institutions worldwide.

Citizen science participation is primarily supported via ***Globe Observer app*** (GLOBE, 2023b) that allows submission of observations of cloud cover, mosquito habitats, land cover, trees, and eclipses (temporary tool available during eclipses only) contributing to a global database that scientists and researchers can access for analysis and study. Prior to submitting any observations, Globe Observer contributors undergo in-app training to ensure accurate and reliable measurements.

To contribute observations and data to all GLOBE Protocols (KSA, 2023) and GLOBE Protocol Bundles (GLOBE, 2023c), citizen scientists can complete GLOBE eTraining modules (GLOBE, 2023d).

GLOBE Protocols outline methods for collecting and documenting Earth science data to meet GLOBE quality standards; the resources are grouped under four Earth spheres: Atmosphere, Biosphere, Hydrosphere, and Pedosphere (Soil) (GLOBE, 2023e).

Co-funded by the European Union, Switzerland and the United Kingdom

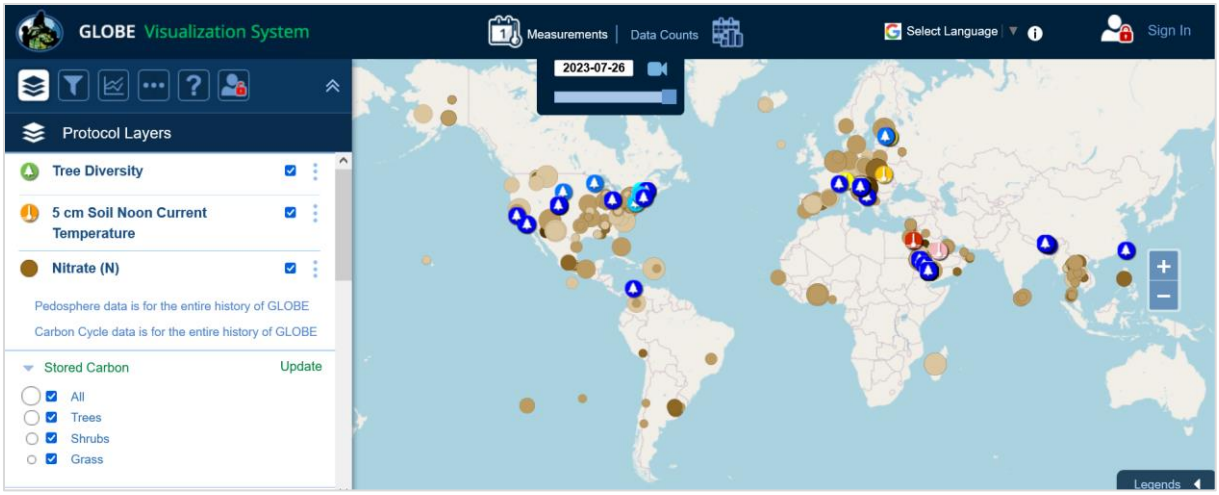


Figure 27: GLOBE Observer Visualisation System.

GLOBE data is freely accessible and dates back to 2015 when the program launched. Data can be accessed via the GLOBE Visualisation System (GLOBE, 2023f) and downloaded using the GLOBE Advanced Data Access Tool (GLOBE, 2023g) or the GLOBE Swagger API (GLOBE, 2023h). Advance Data Access Tool can be used to apply filters (e.g., protocols, date range, data count range, elevation range, latitude and longitude) and download data in CSV format. Swagger API requests return data in either GeoJSON or GLOBE JSON formats.

Selected pre-packaged Globe datasets are available for clouds, dust, eclipse events, land cover and mosquito habitats (GLOBE, 2023i).

4 TOOLS THAT SUPPORT CITIZEN SCIENCE

To achieve data FAIRness, projects must follow good practices 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 the challenge, free and open source tools typically offer several but not all functions to deliver a project end-to-end. For instance, the primary role of Zooniverse (discussed below) is project hosting with additional facilities for basic project search and data storage (for active projects), but it falls short in data hosting and sharing. In this section, we discuss some examples of tools, resources and standards applicable to the citizen science domain. A summary is also offered in ANNEX IV.

4.1 INFORMATION RESOURCES

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. In this section, we discuss ***CitieS-Health Toolkit*** as an example of an information resource designed for citizen science projects.

Co-funded by the European Union, Switzerland and the United Kingdom

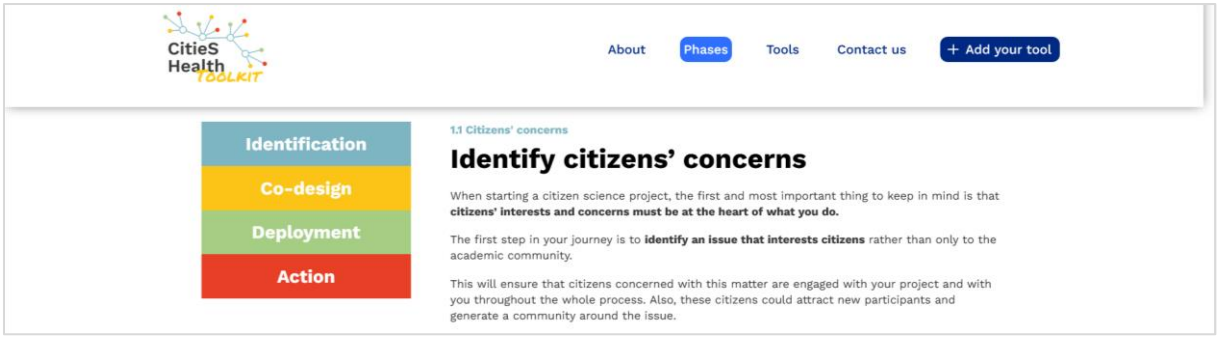


Figure 28: CitieS-Health Toolkit.

CitieS-Health Toolkit is a collection of resources and tools to support individual citizens, communities, scientists, public authorities, and other organisations in defining, designing, deploying, and disseminating participatory studies that tackle environmental and health issues (CitieS-Health, 2023). The platform is currently managed by Ideas For Change company (Ideas for Change, 2023) on behalf of the CitieS-Health project. CitieS-Health project received funding from the European Union's Horizon 2020 research and innovation programme between January 2019 and June 2022.

The toolkit resources are organised around four main stages of participatory studies.

Identification: In this initial stage it is important to identify an issue that concerns or interests citizens and is not only relevant to scientific community. Resources that fall under this topic focus on identifying important issues, defining research questions, and building a community.

Co-design: This phase is concerned with designing the study protocol; it is advised to solicit scientific community to ensure that the study results are of high quality. The resources in this category cover study protocols, instruments for collecting data, and project governance.

Deployment: The resources that belong to this category address participants recruitment and engagement, data analysis and interpretation of the results, and the assessment of impact.

Action: Once the participatory study is completed, it is important to disseminate the results to reach relevant stakeholders such as citizens, scientific community, authorities, industries. Each of these stakeholder groups will require different channels of communication. The resources in this category help with dissemination of the results and reaching different stakeholders, formulating actions based on the results, and project legacy.

At the time of writing, the toolkit contains 31 resources. New resources can be submitted and published on the platform under the creative common licence CC BY-SA 4.0, subject to review and approval.

4.2 PLANNING AND 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 and governance, data hosting, and data sharing.

Many citizen science platforms offer information resources and training materials for designing data management plans. EU-Citizen.Science, for instance, offers training courses including *FAIR Data in Citizen Science Projects* and *Doing Citizen Science as Open Science*. 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 manually inspect resources that appear relevant. Some examples of information resources related to data FAIRness are presented in ANNEX III.

Co-funded by the European Union, Switzerland and the United Kingdom

4.3 PROJECT HOSTING

Citizen science projects can either be hosted on a dedicated third-party platform or 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 (e.g., portal for participants, advance data collection, project stats). Here we discuss examples of citizen science project hosting platforms – ***CitSci 2.0***, ***Zooniverse***, ***Anecdata***, and ***FieldScope***.

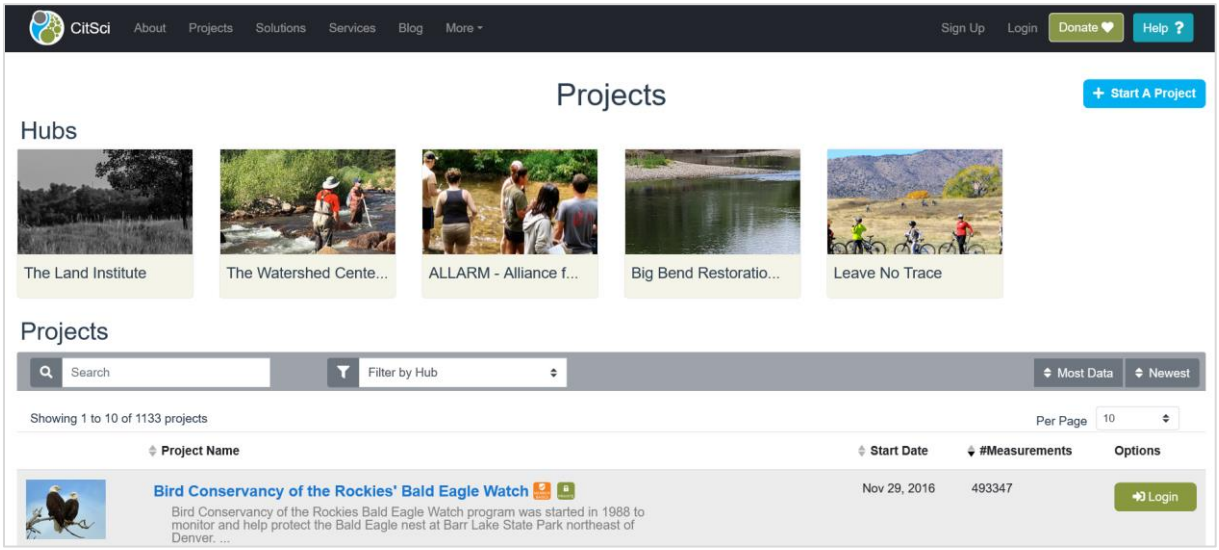


Figure 29: CitSci 2.0 project hosting platform.

CitSci 2.0 was established in 2007 by a group of researchers studying invasive plants in national forests in the US (CitSci, 2023). The platform is supported by the Natural Resource Ecology Lab (NREL) at Colorado State University. It facilitates building and hosting of citizen science projects from across the globe, although most projects hosted are US-based.

The platform offers free tools to facilitate 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 (CitSci, 2023). Project observations can be added via a webform or CitSci mobile app. Only project members can contribute data, memberships can be open (any registered users can join without owner approval) or closed (requests to join require owner approval).

CitSci 2.0 developers API provides facilities to create projects, register users, create datasheets, submit observations, view projects, view observations, and view locations.

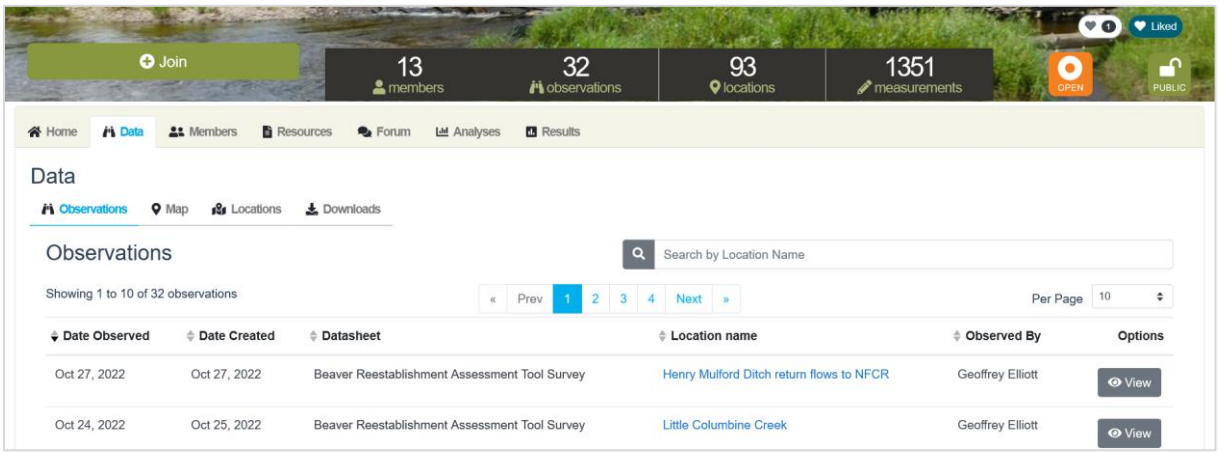


Figure 30: CitSci platform – example public project with data access (Colorado Headwater BRAT Survey).

Co-funded by the European Union, Switzerland and the United Kingdom

The platform does not support advanced project discovery beyond searching by keywords and filtering by project hubs (groups of related projects). Projects created and hosted on CitSci 2.0 can be automatically published to SciStarter for discovery by potential participants and other interested stakeholders which facilitates federation of information. At present, CitSci integrates with SciStarter via project sharing and event tracking. Integration with Zooniverse, iNaturalist and CyberTracker are in the development pipeline.

Project owners can permit project members or the public to view data in a tabular format or on a map and download data in Excel or CSV format (Figure 30). Most projects choose to restrict data download for members only or entirely disallow downloads. The platform hosts 1,133 projects and has 147,504 observations.

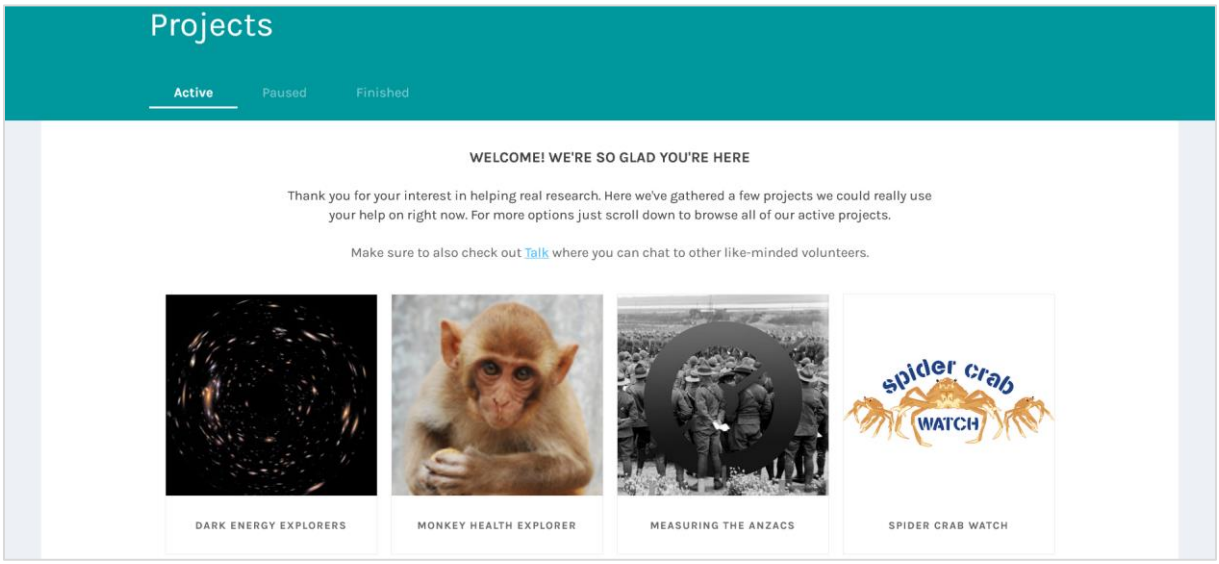


Figure 31: Zooniverse project hosting platform.

Zooniverse is an online platform for hosting citizen science projects in various domains, including arts, biology, climate, history, language, and many more (Zooniverse, 2023). The platform was launched in 2007 (former Galaxy Zoo) by a team of researchers from the University of Oxford and the Adler Planetarium; it is now owned and operated by the Citizen Science Alliance.

The primary objective of Zooniverse is to engage the public in scientific research and leverage the collective power of volunteers to analyse and interpret large datasets that would be challenging or impractical for researchers, otherwise. Volunteers can participate in tasks such as classifying or annotating images, transcribing historical documents, identifying patterns in data, etc.

At present, Zooniverse lists 97 active 243 paused, and 110 finished projects. The platform only supports searching for projects by domain or project name. While the platform does not support sharing of project data, completed projects can publish aggregated results and reports.

Co-funded by the European Union, Switzerland and the United Kingdom

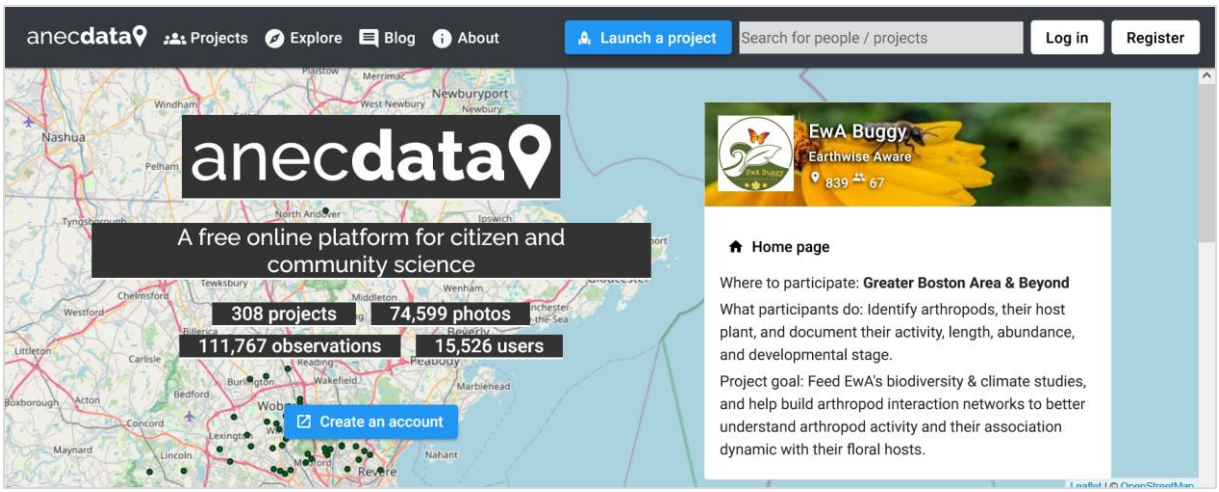


Figure 32: Anecdata citizen science project hosting platform.

Anecdata is a free online citizen science project hosting platform. The platform was founded in 2014 by the Community Lab at the MDI Biological Laboratory in Bar Harbor, Maine (Anecdata, 2023). 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 to submit duplicate observations (e.g., multiple measurements of water quality), define participation mode (open, by request, or closed), 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 and also supports geoprivacy.

The project owners have an option to select either Creative Commons Attribution 4.0 International License or Open Data Commons Attribution License (ODC-By) v1.0 for the data collected via Anecdata platform.

The platform facilities project discovery by project name, postcode, project status, SciStarter affiliation, and project tags. Anecdata additionally supports integration with SciStarter so projects hosted on Anecdata can be made discoverable via SciStarter discovery platform. Similar to CitSci, project membership for contribution can be open (any registered users can join without owner approval) or closed (requests to join require owner approval). Project owners can allow registered users to download project data in CSV format, where the project data file also includes relative URLs to the observation images. Anecdata does not support vocabularies for defining data sheets, hence it is not possible to assume anything about the nature of data content.

Anecdata also 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 (user can draw a bounding box on a map). There is no option to filter observations by tags, which limits the ability to obtain all observations for the required domain or topic. Like public project data, public observations can be downloaded in CSV format; however multiple downloads are required to obtain all available data for a particular domain of interest, since data needs to be filtered by project name to specify the domain.

The platform contains over 300 active projects, 15,500 users, 111,000 observations, and 74,000 photos and mages.

Co-funded by the European Union, Switzerland and the United Kingdom

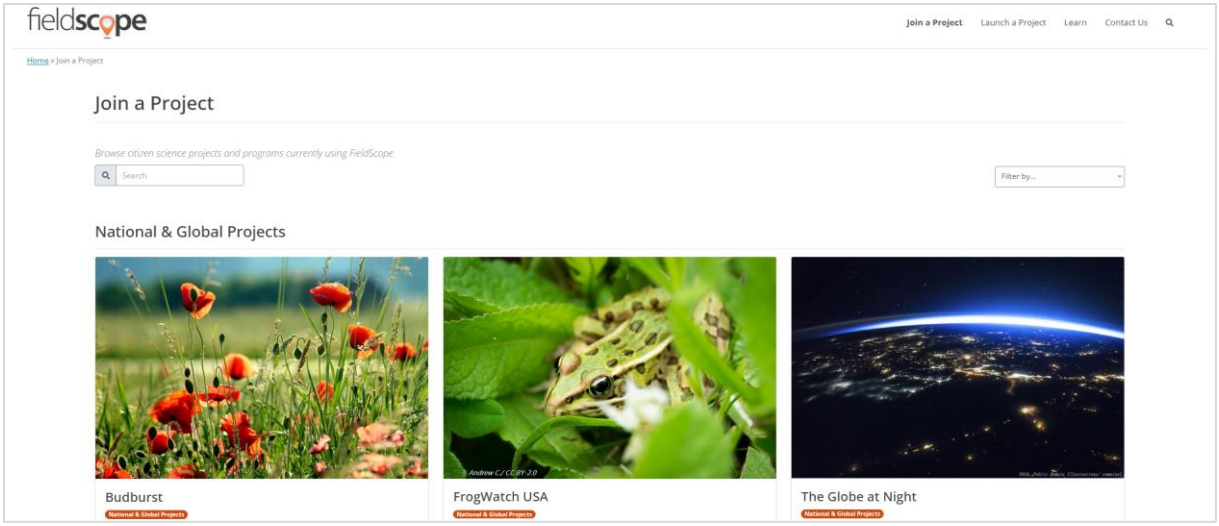


Figure 33: FieldScope project hosting platform.

FieldScope is a citizen science platform that facilitates data collection, map-based data visualisation, and analysis of trends, patterns, and changes over time (FieldScope, 2023). In addition to hosting citizen science projects, FieldScope supports the ***Invitations to Inquiry*** initiative (BSCSa, 2023) where middle and high school students are offered short learning activities to collect, visualise, and analyse large sets of environmental data.

The initial free, web-based version of FieldScope was developed in 2007 by the National Geographic Society in conjunction with the GLOBE Program focused on providing K-12 schools with access to hydrology data. Between 2010 and 2015, with the help of funding from National Science Foundation, FieldScope was transformed into a platform capable of hosting multiple networked projects with private groups of users, data collection forms, data analysis and visualisation tools. In 2015, FieldScope software and the name were acquired by a non-profit organisation BSCS Science Learning (BSCS, 2023b) from the National Geographic Society.

At the time of writing, the platform primarily hosts US projects: 5 national and global projects, and 24 regional US projects.

To host a new or migrate an existing project, the interested party needs to make an enquiry to the FieldScope support team. With the help of various grants, onboarding of new citizen and community science projects costs $50 US dollars until 2025. In general, FieldScope licensing fee varies from $50 (<100 observations) to $5,000 (>10,000 observations) US dollars per annum depending on the number of new observations contributed per year. FieldScope is actively seeking additional funds to ensure low or no cost for hosting the projects. It is free for users to participate, but interested users need to apply to join a project and receive project application approval.

For hosted projects, FieldScope supports migration of project data using spreadsheets, custom datasheets, unlimited number of media files upload (up to 5mb image, and 100mb audio/video files), monitoring of participation, group project management, visualisation of data on a map, graphs and charts, and data download.

The platform does not provide a mobile app but supports a mobile-friendly web interface for data collection. Project observations can be exposed as a table or displayed on a map, but public data sharing and download is not currently supported.

Co-funded by the European Union, Switzerland and the United Kingdom

Table 4: Summary of suitability of project hosting platforms for different citizen science project aims.

Platform	Suitable for	Poor fit
CitSci 2.0	Running multiple projects hosted by same organisation. Complex data collection protocols. Question-first participatory science. Collection of presence/absence species data and other species measurements Collection of data and classification via Zooniverse platform. State-wide programs in the US. Water quality monitoring projects. Assigning volunteers to different sites. Using own Open API to build own apps and embed content within own website.	Checklists Presence only biodiversity data (iNaturalist) Need for AI species identification tools (iNaturalist) Anonymous participation (Anecdata) Image classification only (Zooniverse) Classroom focus, specific protocols (Globe) Educational projects (FieldScope, Globe) Children under 13
Zooniverse	Large project datasets. Assistance with volunteer recruitment. Projects with limited development budget. Projects requiring hybrid human and AI/ML efforts. NASA research projects (NASA-Zooniverse partnership).	Data collection projects Highly custom interfaces or tools Sensitive data Large-scale restricted participation
Anecdata	More complex biodiversity protocols such as recording absence data. Water quality monitoring. Litter recording and clean up. Collection of non-biodiversity image observations.	Presence only biodiversity data (iNaturalist) Image classification only (Zooniverse)
FieldScope	Map-based data analysis and data visualisation. Educational projects. Having multiple groups within the project.	Need for AI species identification tools (iNaturalist) Image classification only (Zooniverse)

4.4 DATA COLLECTION

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, management and retrieval of observations and metadata and the implementation of secure IoT protocols become essential. Enriching the data will involve annotation, classification, or analysis / validation workflows. Here, we discuss examples of open source (***ODK***, ***Epicollect5***, ***Pybossa***) and commercial (***SPOTTERON***, ***ArcGIS***) data collection tools.

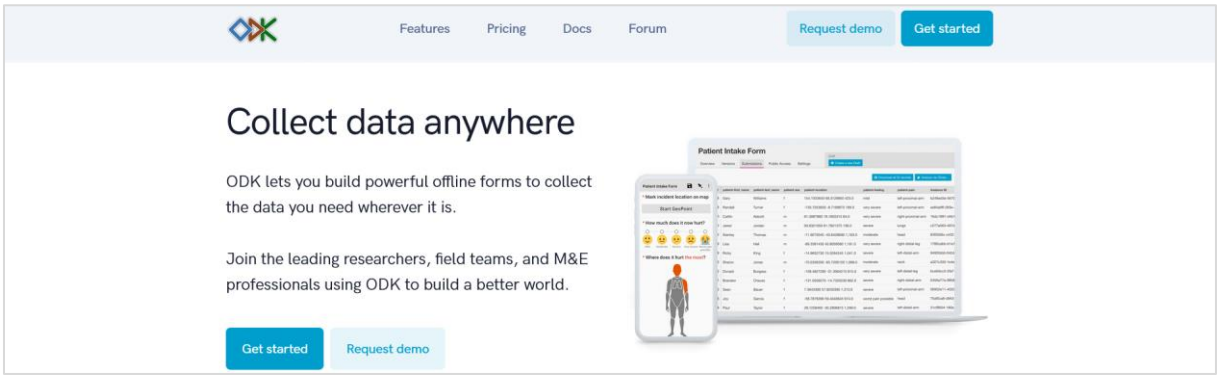


Figure 34: Open Data Kit data collection platform.

Open Data Kit (ODK) is a free and open source set of tools for building custom data collection forms on mobile devices (ODK, 2023). ODK was initially developed in 2008 as a Google.org sponsored project by the University of Washington's Computer Science & Engineering department as a tool for gathering data in resource-constrained environments. The platform is now widely used worldwide, particularly in the fields of public health, humanitarian aid, environmental monitoring, and social research.

Co-funded by the European Union, Switzerland and the United Kingdom

ODK provides a flexible and extensible framework that can be customised and integrated with other tools and systems. It enables efficient and reliable data collection, particularly in offline or low-connectivity environments. Three key components of the ODK platform are ODK Collect, ODK Build, and ODK Central (replacement for ODK Aggregate and ODK Briefcase).

ODK Collect is an Android application for building custom data collection forms and capturing data in various formats including text, images, audio, video, free-hand drawings, barcodes, GPS coordinates, etc. ODK app allows users to download and fill out forms offline; the collected data can be saved locally on the device and later uploaded to a server or cloud storage.

ODK Build is a no-code web-based survey designer tool that enables users to create customise forms. It is best suited for designing simple data collection forms. ODK Build offers a drag-and-drop interface for defining question types, specifying skip patterns, and setting validation rules.

ODK Central is the ODK server that acts as a central repository for storing and managing form definitions, user accounts and permissions, and collected data. It allows data collection clients (build using SDK Collect or external apps) to download forms and upload forms data.

ODK for developers offers comprehensive step-by-step documentation on hosting ODK Central, building ODK web and mobile forms, and collecting and managing responses. ODK code is free to download and deploy on a local server or a cloud but requires technical programming skills. ODK additionally offers ***ODK Cloud*** product which is a paid cloud-based version of the software. ODK Cloud uses the same open source code but the platform is fully maintained and managed by the ODK creators.

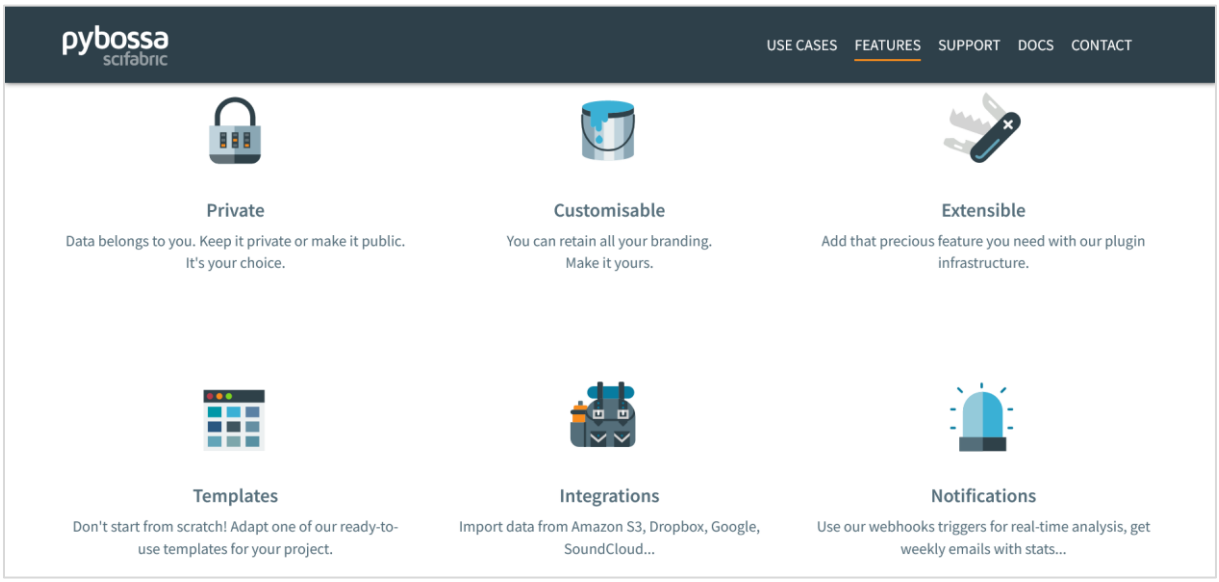


Figure 35: Pybossa crowdsourcing tool.

Pybossa is an open source platform designed for crowdsourcing and citizen science projects that need to analyse or enrich data (Pybossa, 2023). It supports tasks, such as data annotation, image classification, text transcription, sentiment analysis, and more. The free version requires extensive technical skills, but the platform also provides paid subscription plans. ***CS Project Builder*** (Citizen Science Zürich, 2023) offers a no-code open source crowdsourcing platform based on Pybossa framework and designed for data-analysis citizen science projects.

Co-funded by the European Union, Switzerland and the United Kingdom

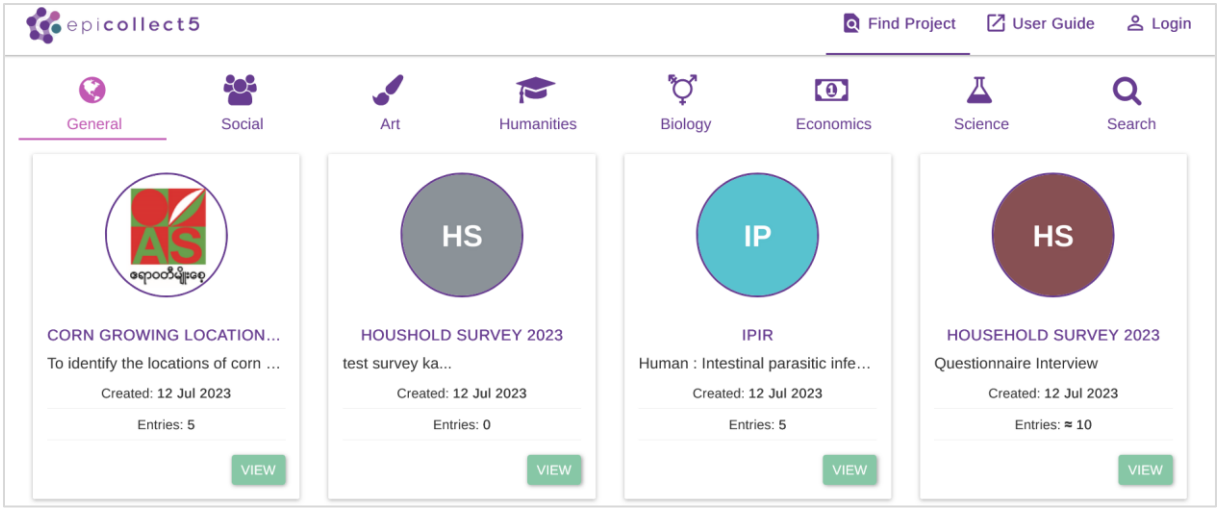


Figure 36: Epicollect5 data collection platform.

Epicollect5 is a free open source data collection platform (Epicollect5, 2023). The platform's primary function is mobile data-gathering in a broad sense, rather than being exclusively tailored to citizen science initiatives. Epicollect5 was established in 2008 and is located in the UK. The platform is managed by the Centre for Genomic Pathogen Surveillance (CGPS) and the commercial arm Digital Epidemiology Services LTD (DES) and is financially supported by the Welcome Trust Foundation.

The platform provides facilities for project creation, custom datasheets with media files (uploads are not limited but Epicollect5 is not a data storage platform), Android and iOS mobile apps for data collection, offline data collection mode. Epicollect5 also hosts the projects and supports basic project discovery using keywords and domain filtering. Project data can be publicly shared in a tabular format and as a map visualisation, data can also be downloaded in CSV and JSON format.

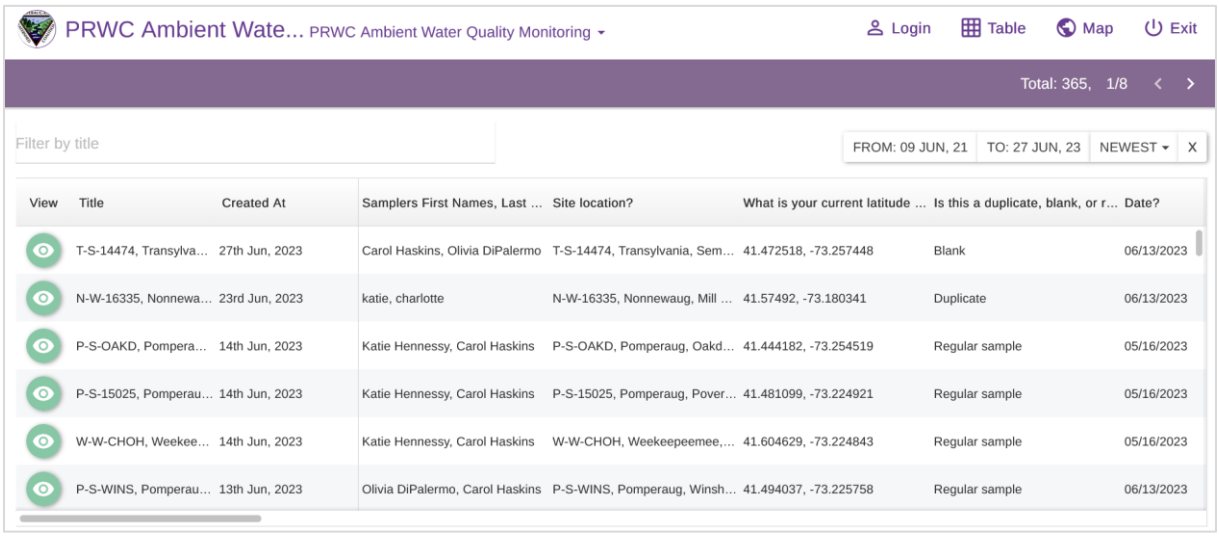


Figure 37: Epicollect5 example data table.

While Epicollect5 supports data sharing for active projects, it is not designed for project hosting or data storage, hence completed projects can move their data at any point in time. Also, it is not possible to filter projects by geo-location or dates so finding relevant data becomes a substantial issue.

Co-funded by the European Union, Switzerland and the United Kingdom

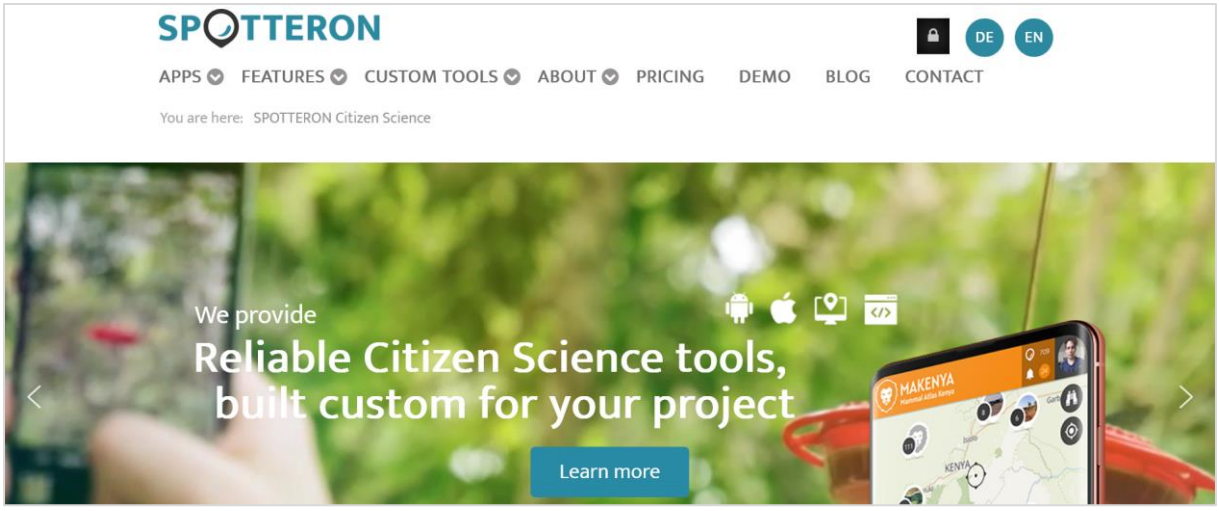


Figure 38: SPOTTERON Citizen Science data collection platform.

SPOTTERON Citizen Science is a web-based platform that facilitates citizen science projects and environmental monitoring initiatives (SPOTTERON, 2023). The platform was established in 2014 and is owned by a commercial company SPOTTERON Gmbh (Austria). The company offers services and partnership to Horizon 2020 and Horizon Europe projects that encompass citizen science components.

SPOTTERON offers a fixed price model for all features in 3 packages: Package L €7.900,- (interactive browser map), Package A €17.900,- (smartphone apps and interactive map), and Package B €20.450,- (apps, interactive map and custom homepage). All subscription plans include 2 years of free regular support, after which support is charged between € 90,- and € 195,- per month.

The platform provides the following features:

- custom fully customisable smartphone apps (Android and iOS);
- data input and observations with rich variety of input fields including autocomplete input/output, spatial input, taxonomy system;
- data management including country or region-based admin access for project partners and public Open Data download;
- project ecosystem including cross-project participation,
- creating communities with advanced functionalities such as newsfeed, comments, liking, following and push messaging;
- data quality with location correctness, data points validation and community validation;
- user motivation with rewards for activity and participation;
- offline mode to contribute observations in low-connectivity or offline conditions;
- interactive maps for observations data;
- fully customisable project website; and
- continuous system maintenance and monitoring.

Figure 39: SPOTTERON Citizen Science features ecosystem.

SPOTTERON platform can act as a data collection, project hosting, data storage, and data sharing system for citizen science projects. The platform is not designed for long-term data preservation or data discovery. While SPOTTERON support Open Science and follow Open Data policies to offer anonymised datasets to the public, it is likely that the projects will move their data after data collection activities are no longer required.

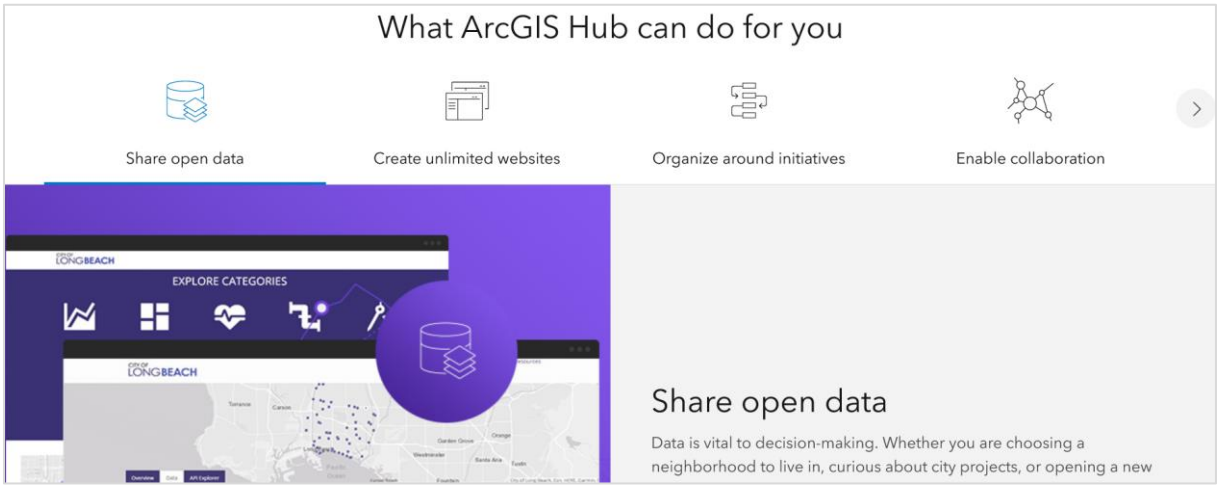


Figure 40: ArcGIS Hub home page.

Esri (Environmental Systems Research Institute) is an American multinational private company that specialises in geographic information system (GIS) software and solutions. It was founded in 1969 Jack and Laura Dangermond and has since become a leading provider of GIS technology. ESRI's flagship product is ***ArcGIS***, a powerful suite of software tools for capturing, managing, analysing, and displaying geospatial data (ArcGIS. 2023a).

While Esri's products are not open source, in recent years, Esri is widening its engagement with the open source community by releasing open source products such as ArcGIS Editor for OpenStreetMap and Esri Geoportal Server and incorporating open source programming languages such as Python.

Esri offers a suit of software products that can be used by citizen science projects from setting up the project and collecting data to processing the results and sharing open data with the community.

ArcGIS Open Data (ArcGIS, 2023b) is a cloud-based platform that facilitates creation of custom sites using existing ArcGIS templates (no coding is required). Open Data capabilities are included in ArcGIS Online

Co-funded by the European Union, Switzerland and the United Kingdom

subscription plan which also covers data storage, advance data analysis, interactive maps creation, and multiuser collaboration.

Using Open Data, citizen science projects can set up own dedicated sites to share their open data. The platform also enables project owners to visualise the project results using interactive maps, graphs or a tabular format. ArcGIS does not support project discovery, only a small sample of ArcGIS-based sites are provided as case studies.

ArcGIS Survey123 is a fully customisable cloud-based survey product. It allows to create custom sophisticated forms for data collection via a web-browser or a mobile app that supports Windows, Android and iOS operating systems (ArcGIS, 2023c). In addition to standard survey platform functionality, Survey 123 data collection forms can include lines and polygons, images and audio files, high-accuracy data captured using an external GNSS receiver, Spike laser measurements, and existing feature services. Survey 123 forms are fully configurable and can support complex data collection needs (e.g., prepopulating answers, automating answers based on date and time, performing calculations).

ArcGIS QuickCapture is a cloud-based survey product designed to capture field observations data from moving vehicles (ArcGIS, 2023d). The license can be purchased as an add-on to the existing suit of ArcGIS products.

QuickCapture offers mobile apps compatible with Windows, Android and iOS operating systems. Captured data can be sent for processing and analysis in real time. The app is fully customisable and facilitates capture of images as well as sensor information from the device. Data collection thresholds can be configured to meet required horizontal accuracy before data is captured.

ArcGIS Community Science Solution is a product specifically designed for collecting location-enabled plant and animal observations from citizen scientists (ArcGIS, 2023e). Community Science is primarily used by conservation organisations, natural resource departments, and other government agencies to engage local communities and collect biodiversity data. Community Science requires ArcGIS Online, ArcGIS Survey123, and Survey123 Connect to be deployed.

ArcGIS Conservation Outreach is a product designed for management of protected areas to monitor wildlife conflicts (ArcGIS, 2023f). This product is primarily used by conservation organisations that monitor conflicts between humans and wildlife. It allows to solicit incident reports, monitor conflict incidents, and manage the resolutions. Conservation Outreach requires ArcGIS Online and ArcGIS Survey123 to be deployed.

ArcGIS StoryMaps is a map-based data visualisation product that allows to create influential interactive maps augmented with text, photos, and videos (ArcGIS, 2023g). StoryMaps builder functionality allows to annotate existing ArcGIS maps, create and annotate new maps, or use a portfolio of templates to design professional-looking content.

4.5 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, sharing, interoperability, (re)use, and integration (especially important 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 their own vocabularies. CitSci, for instance, does not endorse any semantic resources (datasheet templates are under development) which results in unpredictable data structure, ultimately impacting interoperability.

In the following section, we discuss a range of semantic resources and resource collections: while none of these are specific to citizen science, they include many concepts and terms which are very relevant to citizen science initiatives.

Co-funded by the European Union, Switzerland and the United Kingdom

Darwin Core 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*" (TDWG, 2023).

The terms are grouped in nine classes where six describe broad aspects of the biodiversity domain (event, location, geological context, occurrence, taxon, and identification) and two cover relationships to other resources, measurements, and generic information about records (Figure 41).

Record-level Terms	Dublin Core terms, institutions, collections, nature of data record	Simple Darwin Core (flat)
Occurrence	evidence of species in nature, observers, behavior, associated media, references.	
Event	sampling protocols and methods, date, time, field notes	
Location	geography, locality descriptions, spatial data	
Identification	linkage between Taxon and Occurrence	
Taxon	scientific names, vernacular names, names usages, taxon concepts, and the relationships between them	
GeologicalContext	geologic time, chrono-stratigraphy, biostratigraphy, lithostratigraphy	
ResourceRelationship	explicit relationships between identified resources (e.g., one organism to another, taxon to location, etc.)	Generic Darwin Core (relational)
MeasurementOrFact	measurements, facts, characteristics, assertions, references	

Figure 41: Darwin Core classes: Simple Darwin Core (Wieczorek *et al.* 2012).

Since Darwin Core carries two functions, we will continue the discussion in the ***Standards for Structuring and Accessing Data*** section.

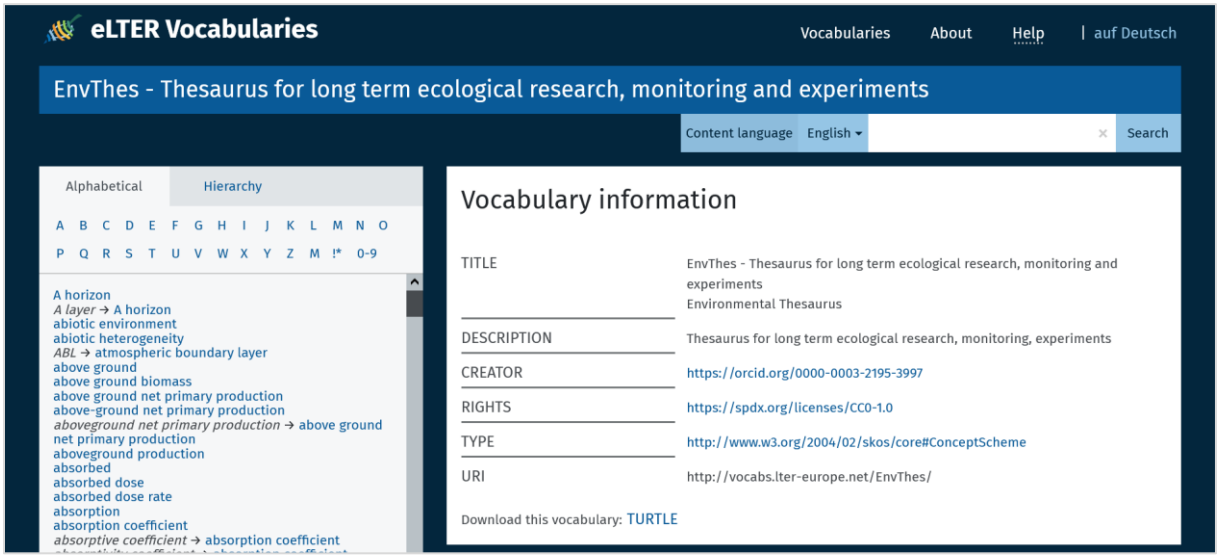


Figure 42: EnvThes (Environmental Thesaurus) – vocabulary information page.

Co-funded by the European Union, Switzerland and the United Kingdom

EnvThes (Environmental Thesaurus) is a controlled vocabulary of semantically well-defined terms that are related to long-term ecological research, monitoring and experiments (eLTER. 2023a). EnvThes was created in 2012 and is managed by LTER-Europe (also eLTER) – the European Long-Term Ecosystem, critical zone and socio-ecological Research.

The backbone of EnvThes is the US-LTER controlled vocabulary, augmented with other existing vocabularies such as NASA Sweet Units, Catalogue of Life, EUNIS Habitats, INSPIRE spatial data themes, and enhanced with additional relevant concepts (Figure 43).

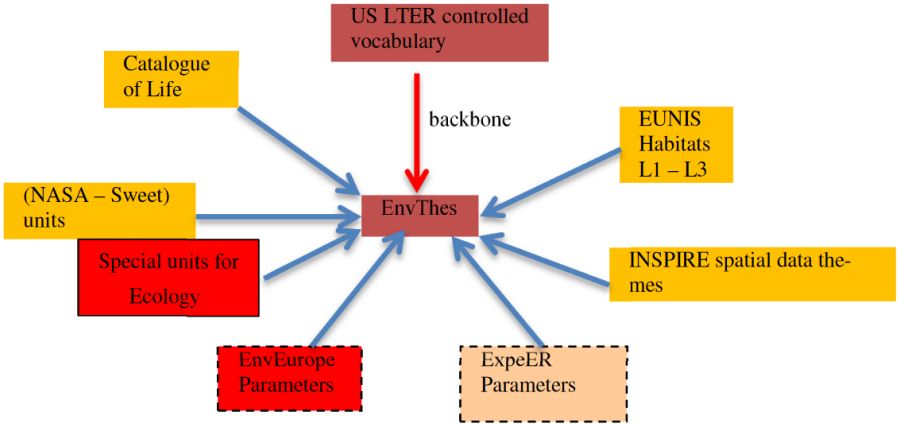


Figure 43: Components of EnvThes (Schentz *et al.*, 2013)

The vocabulary is stored as Linked Data compliant with SKOS model recommendations and can be accessed in human- and machine- readable formats (plain text, RDF/XML, Turtle, JSON-LD). Each term in EnvThes is provided with a definition, relationships to other terms within the vocabulary (broader concept, narrower concepts, exactly matching concepts), alternative terms, notes (e.g., regarding usage and scope), term creator, and exactly matching concepts in external vocabularies. It is distributed under Creative Commons Zero v1.0 Universal (CC0-1.0) license.

eLTER additionally provides the ***eLTER_CL controlled thesaurus*** that contains terms used within the eLTER community (eLTER, 2023b). A ***Standard Observations thesauri*** is currently under development (eLTER, 2023c).

An example consumer of EnvThes is the ***Dynamic Ecological Information Management System (DEIMS-SDR)*** dataset registry for discovery of long-term ecosystem research sites and their data (DEIMS, 2023). EnvThes is also included as part of the EcoPortal semantic resources (EcoPortal, 2023a).

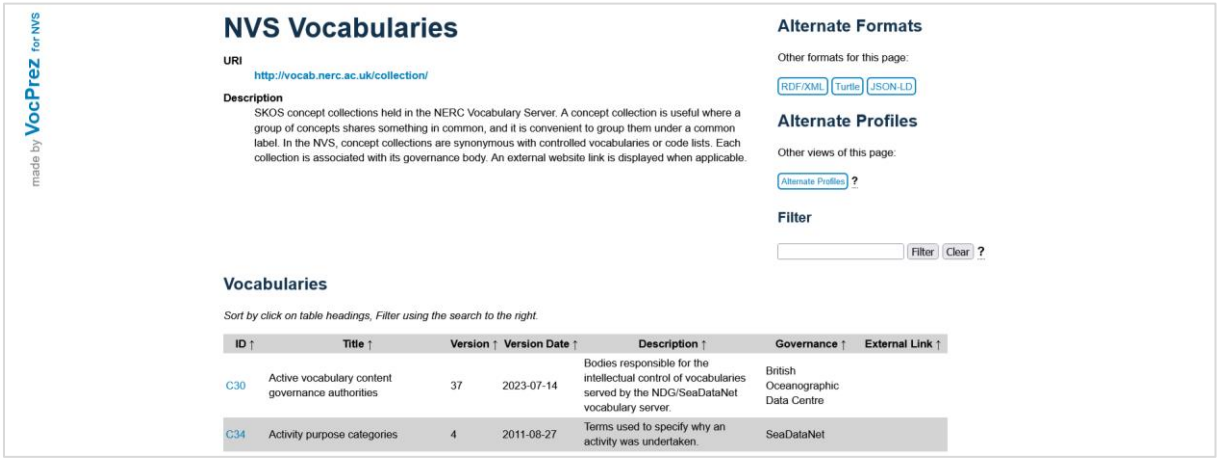


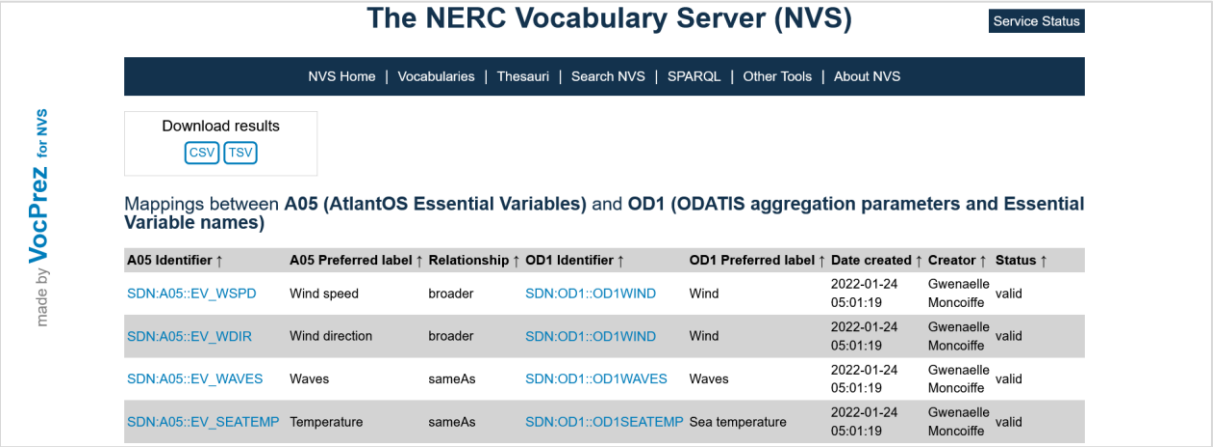
Figure 44: The NERC Vocabulary Server (NVS).

Co-funded by the European Union, Switzerland and the United Kingdom

The NERC Vocabulary Server (NVS) is a collection of standardised and hierarchically structured controlled vocabularies primarily covering the oceanographic and related domains (NERC, 2023a). Version 0 of the NVS has been used to publish controlled vocabularies since 2006. NVS is based on the Simple Knowledge Organization System (SKOS) model and the latest version of NVS – NVS2.0 – was released in in 2012 to align with the updates to SKOS data model recommendations in 2009 (Leadbetter *et al.*, 2012). NVS is funded by the Natural Environment Research Council (NERC) and the platform is managed by the British Oceanographic Data Centre at the National Oceanography Centre (NOC).

The platform comprises vocabularies and thesauri stored as Linked Data compliant with SKOS model recommendations and can be accessed in human- and machine- readable formats (plain text, RDF/XML, Turtle, JSON-LD). The content of the NVS and of these webpages is licensed under CC BY 4.0 (legal code).

NVS supports basic search based on simple text matching, advance search for terms in specified vocabularies or across vocabulary collections, and interrogation of mappings between different vocabularies. An ***Interactive Query UI*** provides a simple interface to query NVS triplestore (RDF database of all NVS vocabularies) using SPARQL queries (NERC, 2023b). This tool also offers functionality to automatically encode SPARQL queries into single line string URLs. Queries can be first tested using the Interactive Query UI for correctness. Query URLs can be decoded into SPARQL queries format.



A05 Identifier ↑	A05 Preferred label ↑	Relationship ↑	OD1 Identifier ↑	OD1 Preferred label ↑	Date created ↑	Creator ↑	Status ↑
SDN:A05::EV_WSPD	Wind speed	broader	SDN:OD1::OD1WIND	Wind	2022-01-24 05:01:19	Gwenaelle Moncoiffe	valid
SDN:A05::EV_WDIR	Wind direction	broader	SDN:OD1::OD1WIND	Wind	2022-01-24 05:01:19	Gwenaelle Moncoiffe	valid
SDN:A05::EV_WAVES	Waves	sameAs	SDN:OD1::OD1WAVES	Waves	2022-01-24 05:01:19	Gwenaelle Moncoiffe	valid
SDN:A05::EV_SEATEMP	Temperature	sameAs	SDN:OD1::OD1SEATEMP	Sea temperature	2022-01-24 05:01:19	Gwenaelle Moncoiffe	valid

Figure 45: Example NVS vocabularies.

Other NVS tools include a ***Vocabulary Editor*** that allows authorities to manage their vocabularies, a ***Vocabulary Builder*** that facilitates building of new terms for review and uptake within BODC Parameter Usage Vocabulary (PUV) semantic model (code name P01), and the ***SeaDataNet P01 Facet Search*** that supports drill-down search for specific terms within the P01 controlled vocabulary (BODC, 2018). The NVS platform also offers a RESTful Swagger API (NERC, 2023c).

Co-funded by the European Union, Switzerland and the United Kingdom

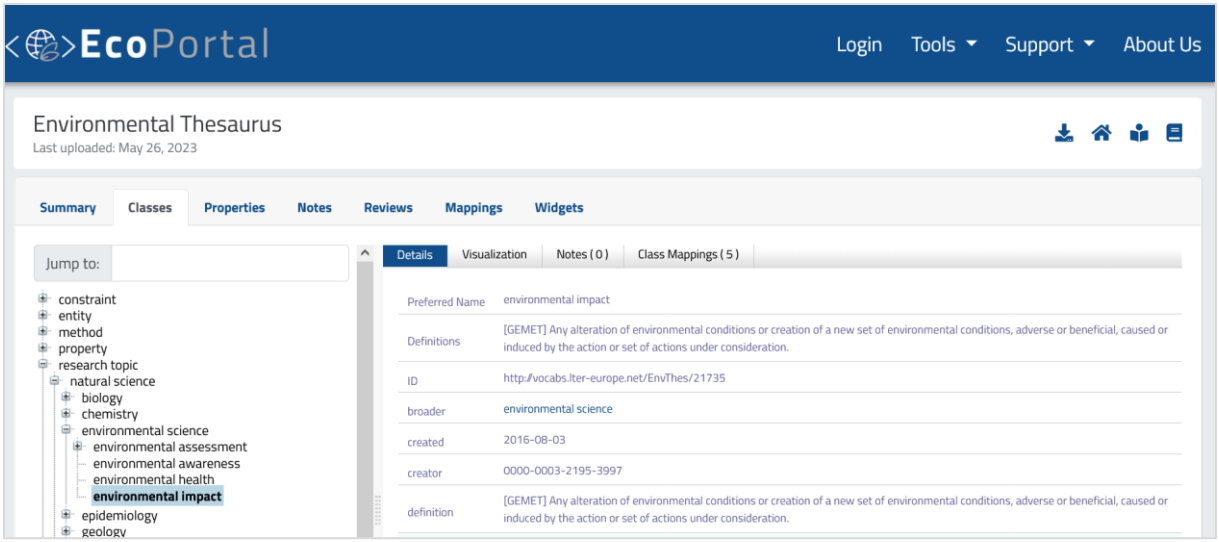


Figure 46: EcoPortal home page.

EcoPortal is a repository of semantic resources and tools for the ecological domain, developed and mainlined by LifeWatch ERIC – the e-Science European infrastructure for biodiversity and ecosystem research (EcoPortal, 2023b). Use of standardised domain vocabularies and ontologies ensures data findability, interoperability and integration. EcoPortal facilitates discovery of controlled vocabularies and core and domain ontologies via user interface or RESTful API. It is designed to support a variety of users including ecology researchers, vocabularies and data managers, ontology experts, and others.

Registered users can submit semantic resources for review to be added to the repository. Resources can be public or private (not indexed in the search), users who submit the resources can specify the visibility of the terms in their semantic resources. At the time of writing, EcoPortal contains 25 semantic resources and 24,184 mappings (associations between terms in different semantic resources).

In addition to semantic resources discovery, EcoPortal offers tools for annotating free text, reviewing mappings between different terms, and get recommendations for semantic resources. These tools are also accessible via EcoPortal RESTful API (EcoPortal, 2023c).

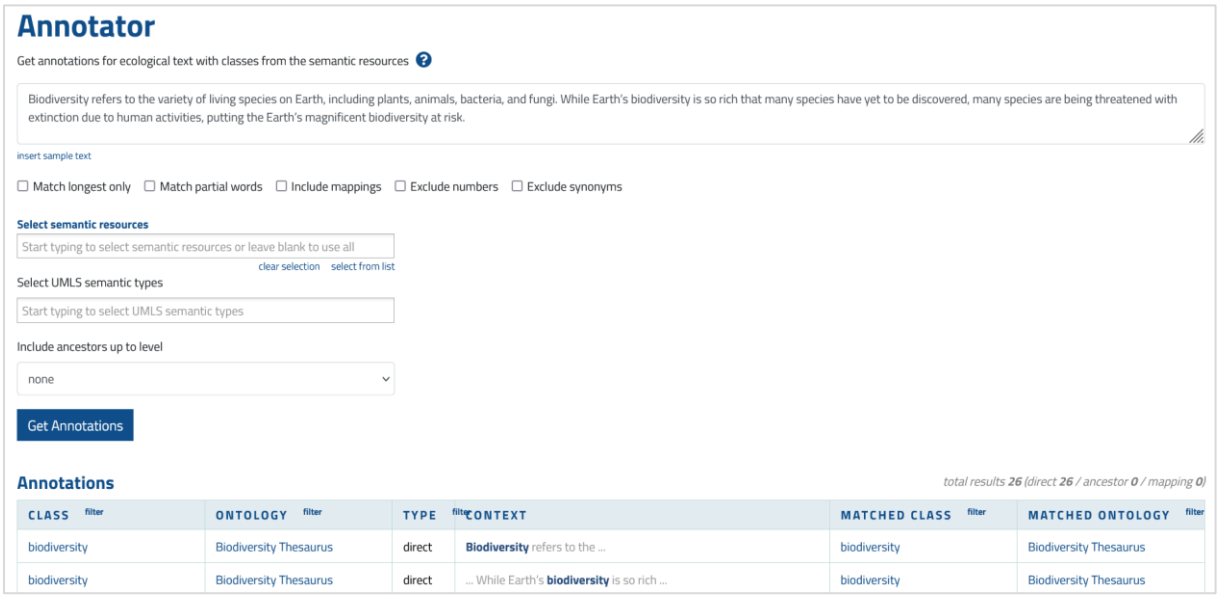


Figure 47: EcoPortal Annotator tool.

EcoPortal Annotator is a tool that allows to annotate free text with the terms contained within EcoPortal semantic resources. The annotator uses the exact string comparison between the text supplied and the semantic resource term names, synonyms, and IDs. The set of matches can be expanded using hierarchical expansion where mapped terms are also included in the search.

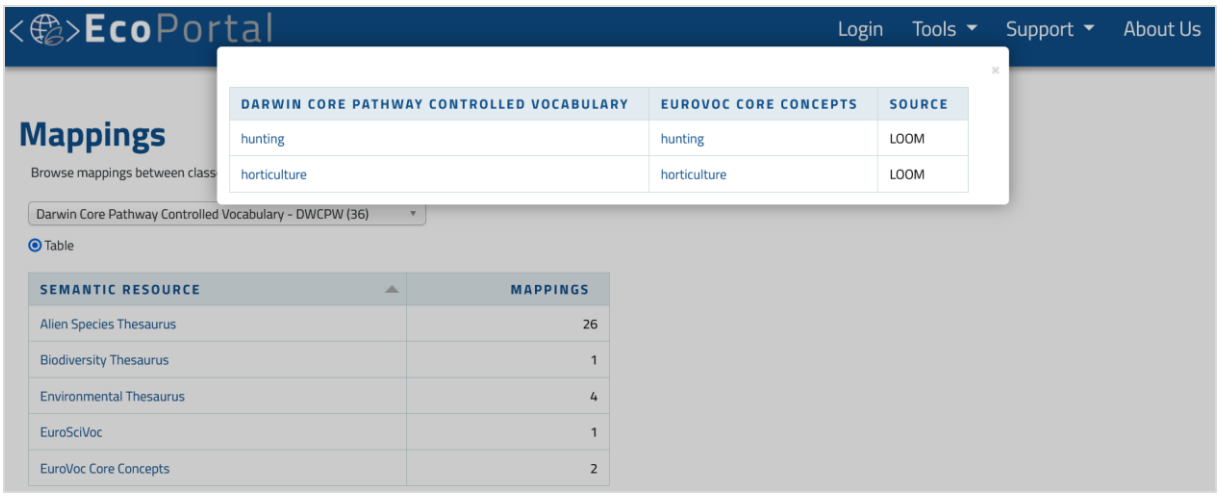


Figure 48: EcoPortal Mappings tool.

The ***EcoPortal Mappings*** tool allows identification of the associations between two or more terms in different semantic resources contained within EcoPortal. To browse semantic resources mappings, the user can navigate to the Mappings tool and select the semantic resource of interest. In most cases, the association represents a degree of similarity between the terms.

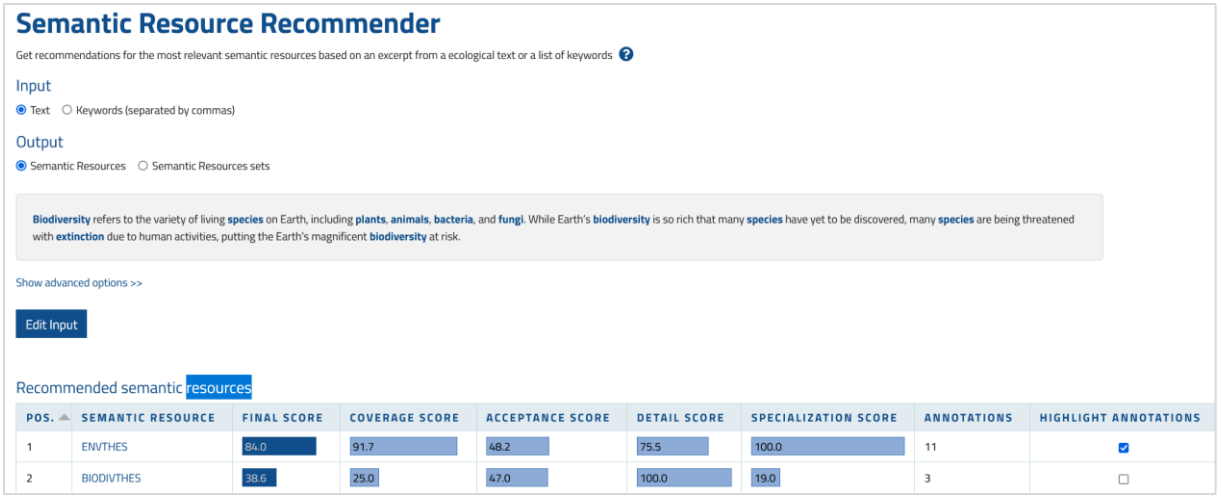


Figure 49: EcoPortal Recommender tool.

The ***EcoPortal Recommender*** allows a user to obtain recommendations for the appropriateness of the semantic resources for an ecological text or a set of keywords. To obtain recommendations, users can navigate to the Recommender tool, enter free text or keywords, adjust the default evaluation criteria, and receive a list of ranked recommendations of the semantic resources contained within EcoPortal.

The ranking of the semantic resources is based on four evaluation criteria: ***coverage*** (calculated based on the annotations for the input; these are produced by evoking the Annotator service), ***acceptance*** (calculated based on the number of visits to the semantic resource and its presence or absence in Unified Medical Language System (UMLS)), ***detail of knowledge*** (calculated using the coverage of the input by the number of definitions, synonyms and properties of the semantic resource classes), and ***specialisation*** (calculated

based on the position of each annotated class in the semantic resource hierarchy, the number and type of the annotations done with the semantic resource, and normalised by the size of the semantic resource)

Full User Guide for EcoPortal is available at: https://www.lifewatch.eu/wp-content/uploads/2021/10/EcoPortal-Documentation.pdf.

4.6 DATA PUBLISHING AND PRESERVATION

To ensure long-term value outside of 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 hard to discover. Indeed, Schade and Tsinaraki (2016) in their survey report on Data Management in Citizen Science Projects reveal that the majority of surveyed projects host their data on a remote server (38%) or a local machine (16%) managed by a project member. It however remains unclear whether this data is catalogued and is discoverable elsewhere. Other projects might collect data suited for contribution to larger initiatives that already provide open data capabilities, e.g., iNaturalist, eBird, BioCollect, NASA Globe, Sensor.Community. However, such projects might find themselves collecting additional valuable data that does not conform to the accepted data structure of their chosen publishing initiative.

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. It could also be useful for platforms or projects that decide to close because their objectives were fulfilled, or they cannot cover the costs of maintaining the infrastructure. Examples of existing open repositories include Zenodo and PANGEA (described below), though both are more suitable for static data snapshots than for the continually-evolving datasets which emerge from many citizen science projects.

Inability to support longer-term or ongoing citizen science projects that seek to continually supply their datasets presents a major limitation. While Zenodo supports dataset updates and DOI versioning, it is impractical to generate an excessive number of versions. PANGEA does not support dataset versioning entirely and there is a delay in the publication and depreciation of erroneous datasets due to PANGEA's rigorous review process.

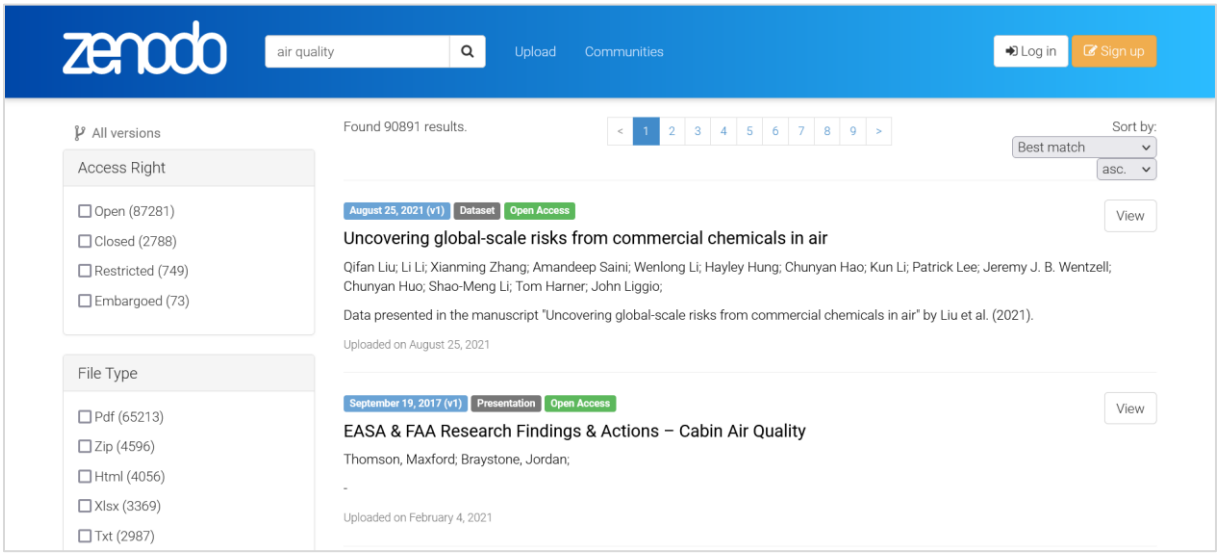


Figure 50: Zenodo data repository platform - example of "air quality" search results.

Zenodo is a multi-disciplinary open repository designed for research communities to deposit research datasets, software, reports, papers, and other digital research artefacts (Zenodo, 2023a). The platform was launched in 2013 and is owned by the European Organization for Nuclear Research (CERN).

Co-funded by the European Union, Switzerland and the United Kingdom

Registered users can deposit research artefacts at any stage of the research lifecycle, provided that they hold appropriate rights for the materials. The maximum allowed file size is 50GB, exceptions may be made for larger files depending on the specific circumstances. Artefacts can be deposited under closed, open, or embargoed access. Zenodo offers RESTful API to support deposit of research outputs, records search, and files upload and download. All uploads are assigned a Digital Object Identifier (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 database enhancing discoverability and accessibility within the open science community (OpenAIRE, 2023); ***CERN Analysis Preservation (CAP)*** infrastructure to enabling researchers to preserve and share their analysis workflows, code, and associated data in a FAIR manner (CERN, 2023).

Zenodo supports advance search by constructing complex text-based search queries and entering these into the search window, e.g., `+title:"open science" publication_date:[2022-01-01 TO 2023-01-01]` (https://help.zenodo.org/guides/search/). However discovering new and relevant geospatially tagged resources and datasets by this means can prove extremely challenging. Unlike arXiv, Zenodo is not yet indexed by Google Scholar.

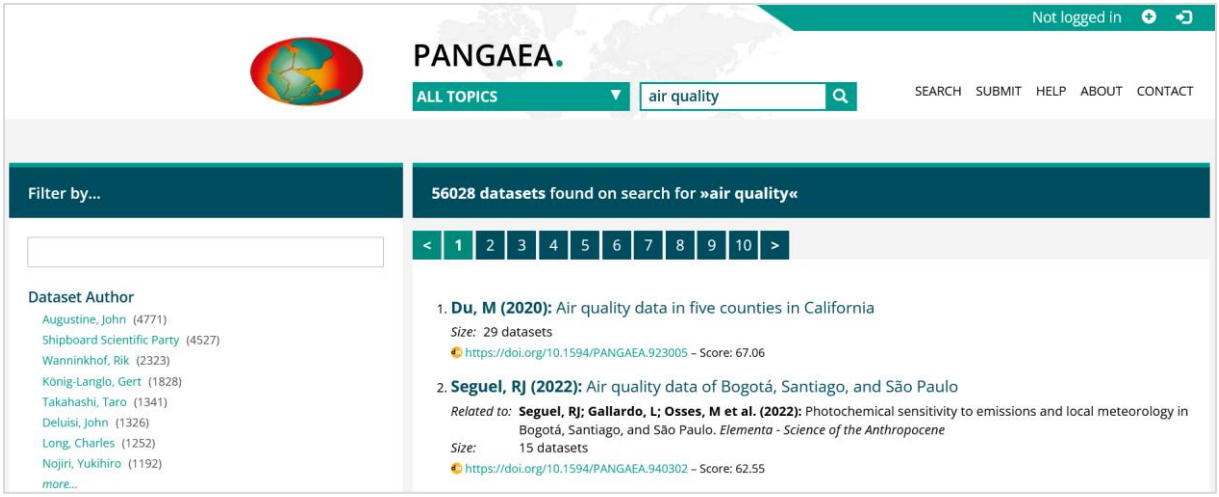


Figure 51: PANGEA data repository platform - example of "air quality" search results.

Publishing Network for Geoscientific and Environmental Data (PANGEA) is an Open Access library and a publishing platform for georeferenced data (PANGEA, 2023a). The platform was established in the 1990s, it is hosted by the Alfred Wegener Institute and is supported by The European Commission, Federal Ministry of Education and Research (BMBF), Deutsche Forschungsgemeinschaft (DFG) and International Ocean Discovery Program (IODP).

Registered users can deposit any data from earth and life sciences domain but each dataset is checked and quality controlled by the PANGEA support team to ensure machine readability and FAIRness of the archived data. The editorial process can take up to several months, depending on the complexity of the dataset submitted. Data can be submitted in any format and is reformatted by the PANGAEA editors to fit the PANGAEA Data Model (PANGEA, 2023b). All published data is then stored for a minimum of 10 years.

Complex queries can be constructed to search datasets by domain, dataset author, dataset publication year, topic, project, basis, method/device, campaign, location, geographical coordinates, and dates. Search results can be viewed as a table or displayed on an interactive map. Datasets can be downloaded as ZIP files

Co-funded by the European Union, Switzerland and the United Kingdom

containing all dataset data in tab-delimited text format, binary objects (for large files), or other ISO compliant formats (e.g. images, films). Community-developed scripting tools can be used for querying metadata and downloading tabular PANGEA datasets (PANGEA, 2023c). PANGAEA also offers a Data Warehouse tool to download aggregated data from different PANGAEA datasets in a single file.

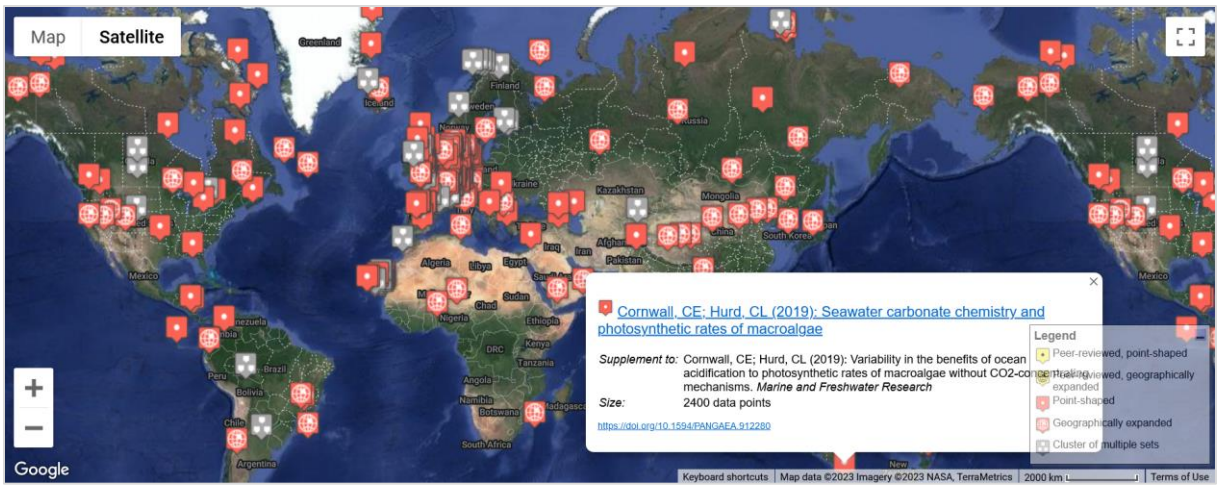


Figure 52: PANGEA data visualisation example.

5 STANDARDS FOR STRUCTURING AND ACCESSING DATA

International data standards ensure consistency and interoperability among data collected by different individuals or groups participating in citizen science initiatives. Their use enhances the credibility and scientific value of citizen science efforts, 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. In this section, we discuss Darwin Core (introduced earlier), the Open Geospatial Consortium (OGC) Observations and Measurements (O&M), 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 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 maintained by the Darwin Core Maintenance Interest Group. Darwin Core is based on Dublin Core, Species Analyst (SpeciesAnalyst, 2001) and the Access to Biological Collections Data (ABCD) standards (Figure 53); it is designed to be minimal data model (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 HTML, RDF/Turtle, RDF/XML, and JSON-LD formats. As discussed previously, Darwin Core is the preferred standard for publishing data to GBIF. Darwin Core is a data model and does not define any specific protocol for exchanging data between users.

Co-funded by the European Union, Switzerland and the United Kingdom

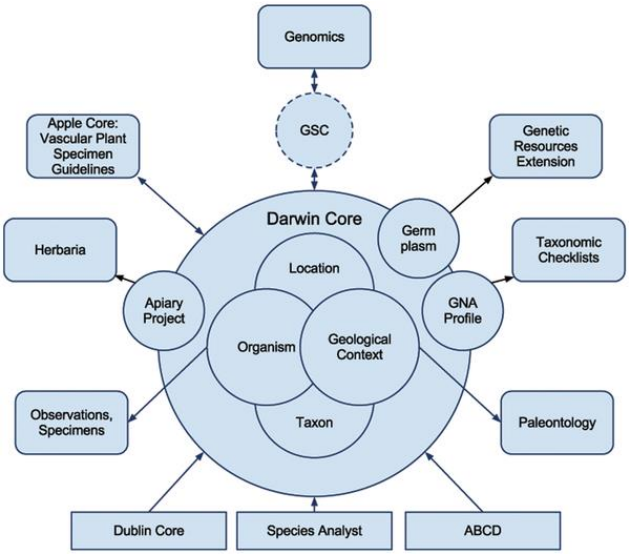


Figure 53: Scope of Darwin Core (Wieczorek *et al.* 2012).

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 data about various types of observations from sensors, instruments, algorithms or process chains. In the context of the O&M conceptual model of the world, citizen scientists could be seen as instruments or sensors themselves that collect observations about a phenomenon.

The O&M data model was initially developed in the context of geographic information systems and is now 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 54): *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 is an example application of the O&M standard at the European infrastructure level (EC, 2015). EMF describes each facility as a spatial object in the context of INSPIRE (DataCove, 2019) and links observations and measurements of environmental parameters to the facility, where citizen science is included as one of stakeholder initiatives for sharing public data. Currently the OGC and the W3C are working on another data model for sensor data that is called Semantic Sensor Network Ontology (SOSA, 2017).

Co-funded by the European Union, Switzerland and the United Kingdom

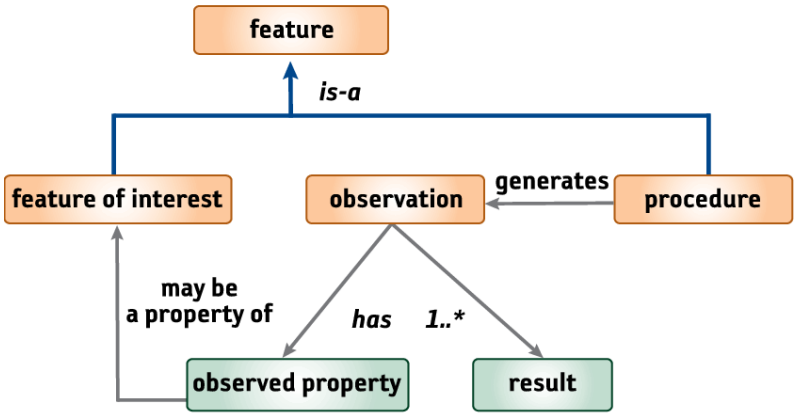


Figure 54: Basic structure of the OGC Observations and Measurement Model (Usländer, Coene and Marchetti, 2012).

OGC SensorThings API 1.1 (STA) provides an open and unified way to interconnect heterogeneous Internet of Things (IoT) devices, data, and applications over the Web (OGC, 2023a). It includes a generic data model for sensor data based on O&M and a communication protocol over HTTP and MQTT based on ODATA. The first version 1.0 was published in 2016 (the latest version 1.1 in 2021), 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 sensing entities of STA are shown in Figure 55. 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). Entities such as *FeatureOfInterest*, *ObservedProperty*, and *Observation* are based on the OGC O&M model.

The increasing use of STA is demonstrated by various IoT-derived platforms for environmental monitoring and smart cities including FROST Server open source implementation of STA (Fraunhofer, 2023), a STA-based INSPIRE download service (EC, 2023e), and the adoption of STA by the French Geological Survey (BRGM, 2023) for the national groundwater monitoring system and water quality database.

Co-funded by the European Union, Switzerland and the United Kingdom

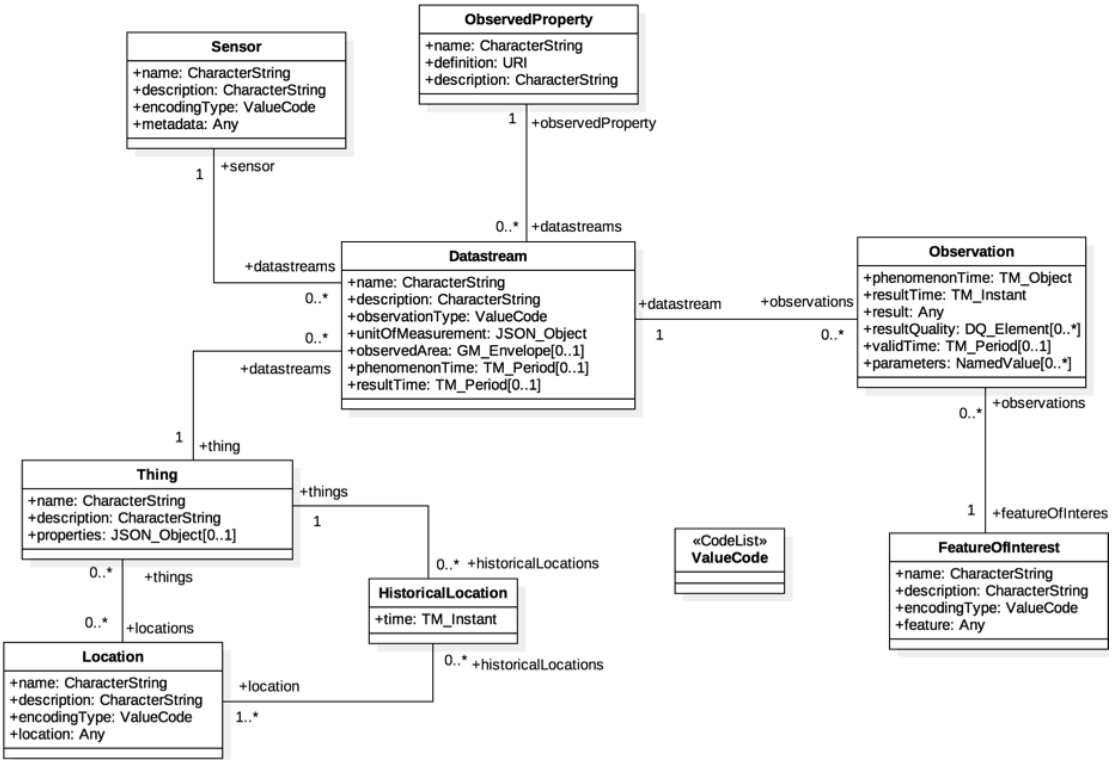


Figure 55: Sensing entities of SensorThings API (OGC, 2021).

OGC SensorThings API Extension: STAplus 1.0 is an approved international standard and is an extension to the STA data model that is based on requirements from the citizen science community (OGC, 2023b). It is designed to support the STA data model where FAIR (in particular, *Interoperable* and *Reusable*) principles can be reinforced by adding entities of ownership, licence and project information for sharing observations. The extension also enables users to express explicit relations between observations as well as create group(s) of observations that belong together.

The STAplus data model describes 5 entities additional to the ones already present in STA: *Party* (links a user to a *Datastream* or *Group*), *License* (specifies reuse conditions), *Campaign* (allows to organise a campaign or project), *ObservationGroup* (allows packaging of individual *Observations* as a bag or set), *Relation* (supports relationships between *Observations*).

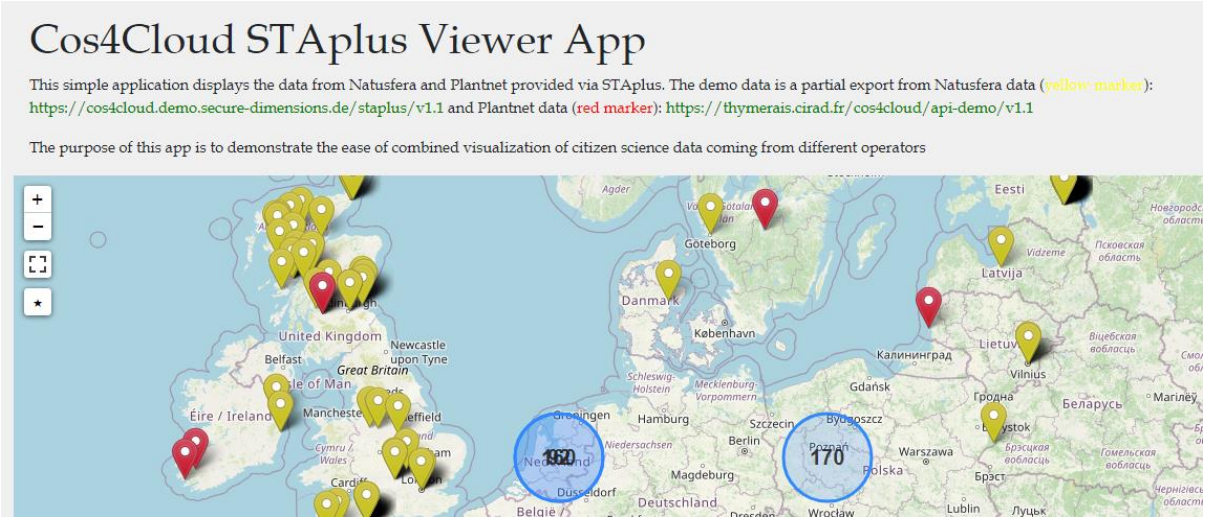


Figure 56: Cos4Cloud STAplus Viewer Application for visualisation of integrated citizen science data.

Co-funded by the European Union, Switzerland and the United Kingdom

Cos4Cloud '*OGC Best Practice for using SensorThings API with Citizen Science*' (OGC, 2022) document offers practical examples of applying the STAplus extension in the citizen science domain.

The camera trap use case offers an example implementation for the companies or projects that wish to purchase camera trap sensors and lend these to citizen scientists for data collection using STAplus. The use case offers walk-through code for the setup of a camera trap, mapping of the different data objects created by the camera trap to the STAplus data model, the use of Batch-processing to reduce communication between the sensor and the STAplus instance, and the relocation of the camera trap sensor to a new site.

The Natusfera/iNaturalist use case offers an example implementation for transforming iNaturalist data into the STAplus data model. Natusfera is a citizen science portal for the iNaturalist data collected in Spain (Natusfera, 2023). The use case provides a schema mapping from the Natusfera data model to the STAplus data model and a JavaScript implementation for reading Natusfera or iNaturalist data (in JSON format) and creating the equivalent STAplus objects.

The "Pl@ntNet data as a Service" use case demonstrates a practical implementation for publishing the Pl@ntNet dataset (which contains over 10 million observations) using a STAplus API. Pl@ntNet is a worldwide citizen science project that uses pictures for the identification of plant species (Pl@ntNet, 2023).

STAplus Viewer App demonstrates how biodiversity data from different citizen science platforms can be aggregated and visualised in a common map viewer (Figure 56) (Cos4Cloud, 2022a).

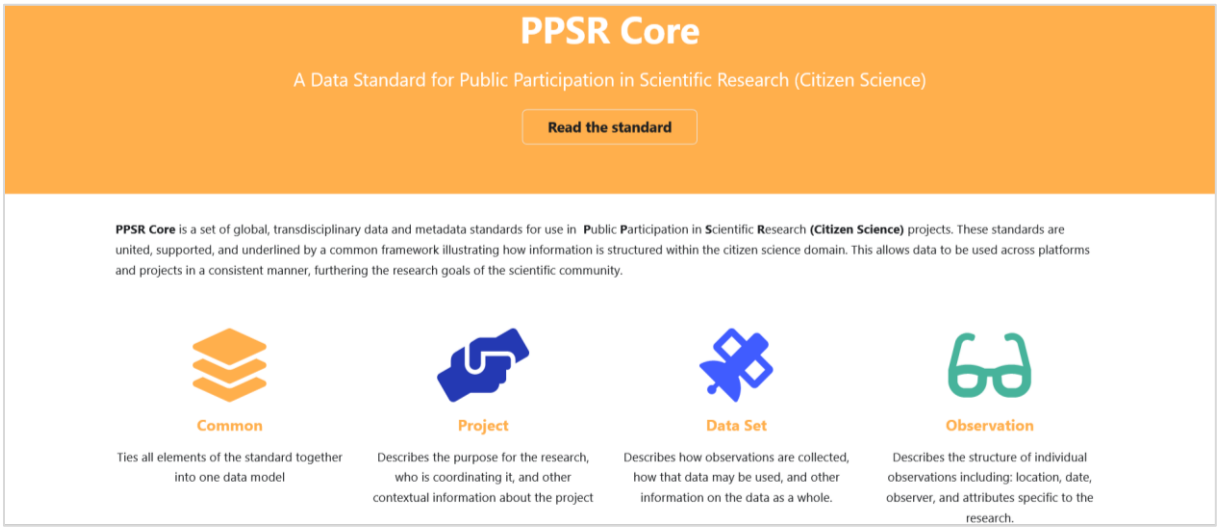


Figure 57: PPSR Core standard home page.

PPSR Core is an open data and metadata standard that defines a common framework for describing citizen science projects and the produced data and observations (CSA, 2021a). The standard is designed to enable sharing of basic common information across databases that catalogue citizen science projects. PPSR Core initiative started in 2013 and was supported by the DataONE PPSR Working Group and SciStarter (representing databases of citizen science projects in SciStarter, CitSci.org, the Cornell Lab of Ornithology, and the Wilson Center). At present, PPSR Core standard is maintained by the Citizen Science Association's Data & Metadata Working Group with the support from volunteers.

PPSR Core standard comprises four models (Figure 57):

The ***Common Data Model (CDM)*** is a data model for aggregating citizen science projects within a common organising framework into programs or campaigns. The model requires complete implementation of Project, Dataset, and Observation models.

The ***Project Metadata Model (PMM)*** is a metadata model for describing the purpose, responsible parties, participation and engagement, and other contextual information for citizen science projects. This model supports project discovery by location, time period, participation task, etc.

Co-funded by the European Union, Switzerland and the United Kingdom

The ***Dataset Metadata Model (DMM)*** is a metadata model for describing collections of observations. The model enables dataset discovery by protocols used, temporal rage, licence, quality assurance methods, etc. and helps to make an informed decision about the dataset's fitness-for-use.

The ***Observation Data Model (ODM)*** is a data model for defining domain "profiles", i.e., core sets of features that should be collected for a given study. The model is designed to support datasets integration by specifying domain-specific data standards or data vocabularies.

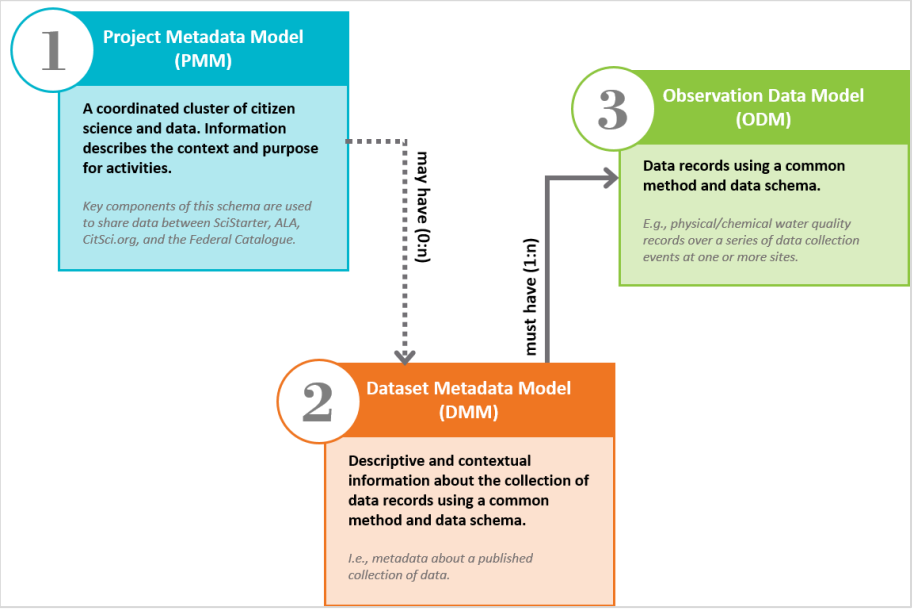


Figure 58: PPSR Core Common Data Model (CDM).

At present, PPSR Core is still under review and development with the ***Project Metadata Model*** being most well-defined and adopted by SciStarter.org, CitSci.org, Atlas of Living Australia BioCollect, and CitizenScience.gov. The ***Project Metadata Model*** includes controlled vocabularies (subject to changes), but projects are welcome to adopt other vocabularies as long as these are clearly referenced.

The ***Observation Data Model*** includes some examples of data standards (Darwin Core for biodiversity projects) but seeks to provide a more comprehensive list. Future development work will focus on gathering the existing domain-specific data standards and vocabularies, developing new models where these do not exist, and identifying common attributes in different domain models to support multi-disciplinary data aggregation.

The PPSR Core standard is available on GITHub (CSA, 2021b) where community is welcome to make contribution or propose changes. The repository also includes a JSON Schema (implementation of the standard) and guidelines on transforming the existing database of citizen science projects into a complaint JSON file.

6 SOME CITIZEN SCIENCE PROJECTS

In the following section, we discuss a range of larger-scale ongoing and past European funded projects and their deliverables that applied citizen science to collect new or validate existing data, engaged in citizen science initiatives, and developed tools and guidelines to support citizen science.

Co-funded by the European Union, Switzerland and the United Kingdom

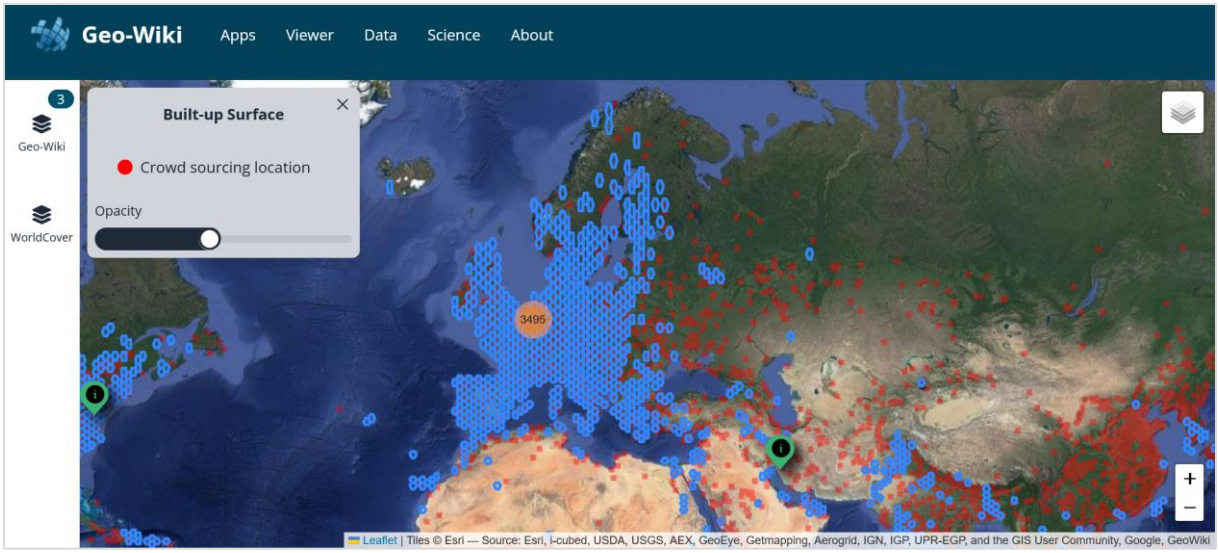


Figure 59: Geo-Wiki data viewer page.

Geo-Wiki is an ongoing collaborative online platform and citizen science project that aims to improve the accuracy of global land cover and land use information (Geo-Wiki, 2023). It was developed to engage the public, scientists, and experts in the task of validating and improving existing land cover maps and datasets. The project was initiated in 2009 by the International Institute for Applied Systems Analysis (IIASA) in Austria.

Geo-Wiki provides a set of apps to enable citizens to participate in the project, some examples are highlighted here. ***LACO-Wiki*** is a web-based application that uses satellite and aerial imagery from Google, Bing and OpenStreetMap for validating land cover and land use maps. ***AgroTutor*** is a mobile application (iOS and Android) designed for agricultural producers to keep a registry of their plots and crops, as well as all agronomic activities. ***Camaliot*** is a mobile application (Android only) designed for collecting Global Navigation Satellite System (GNSS) data to improve weather forecasting models. ***City Oases*** is a mobile application (iOS and Android) designed for documenting the usage of open urban spaces. ***Crowd AI*** is a cross-platform web-based application for engaging citizens in training AI models for detecting tropical deforestation. ***Natura Alert*** is a mobile application (iOS and Android) that enables recording of the damage or loss of sites vital for the survival of bird populations. ***FotoQuest Go*** is a mobile application (iOS and Android) that directs participants to specific locations to track changes in European landscape. ***Yoma*** is a mobile application (Android only) designed for documenting access to potable and clean water.

Geo-Wiki platform facilitates users to view satellite imagery and compare it with existing land cover maps to identify discrepancies or inaccuracies and provide feedback on specific areas where they believe the land cover information is incorrect or outdated. The platform aggregates and analyses the feedback and contributions from users to generate more accurate and up-to-date land cover data. The validated data generated through Geo-Wiki can then be used by researchers, policymakers, and organisations for various purposes, such as assessing deforestation, tracking land use changes, and understanding the impact of land cover on ecosystems and climate.

Geo-Wiki datasets are available on Zenodo platform and can be used under the Creative Commons Attribution 4.0 International license (Zenodo, 2023b).

Co-funded by the European Union, Switzerland and the United Kingdom

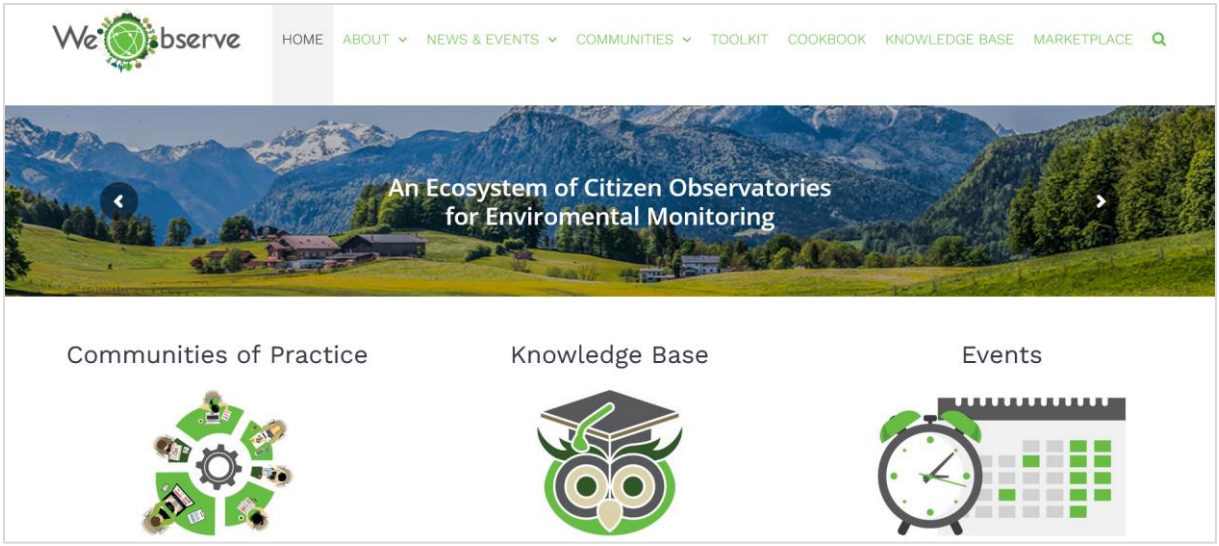


Figure 60: WeObserve project home page.

WeObserve is an EU Horizon 2020 project that ran between December 2017 and January 2021 (WeObserve, 2023). It focused on empowering citizens and communities to actively participate in data collection and environmental monitoring, thus contributing to a better understanding of environmental issues and the development of sustainable solutions. The project worked towards improving the coordination between Citizen Observatories and related regional, European and International activities. Citizen Observatories are defined as "*any use of Earth observation technology in which citizens collect data and are empowered by the information generated from these data to participate in environmental management*" (Grainger, 2017).

WeObserve offers a toolkit of open access tools developed by the project and project partners, namely GROW, SCENT, GroundTruth 2.0 and LANDSENSE (described below). The toolkit is divided into four categories: (1) *Co-designing / co-creating your observatory*, (2) *Training & data collection for environmental monitoring*, (3) *Data quality and visualisation*, and (4) *Evaluation and advocacy*.

GROW is an EU Horizon 2020 Citizen Observatory project that ran between November 2016 and October 2019, at present continues through voluntary community groups (GROW, 2019). The project focused on empowering citizens and community in taking action on soil and climate issues within Europe. One of the key project outcomes was validation of soil moisture data from satellites, including high-resolution Sentinel-1 data, using crowdsourced ground observations from low-cost sensors. GROW supported 24 communities in 13 European countries to create a network of 6,502 ground-based soil sensors resulting in a dataset with 516M rows of soil data.

SCENT is an EU Horizon 2020 Citizen Observatory project that ran between September 2016 and August 2019 (SCENT, 2019). The project aim was to engage citizens in the monitoring of land cover and use changes to support the policy makers. Some of the outcomes of the project include SCENT Explore mobile app for capturing environmental related information (e.g., water level and flow velocity, flood related events), SCENT Measure mobile app that connects to a portable smart sensor to measure soil conditions, and co-design activities and SCENT toolbox evaluation reports.

GroundTruth 2.0 is an EU Horizon 2020 Citizen Observatory project that ran between September 2016 and December 2019 (GroundTruth 2.0, 2019). The project covered the thematic areas of flora and fauna, water availability, and water quality for land and natural resources management. Citizen data was collected via mobile apps and social media analytics. The main achievements of the project include (a) set up and validation of six citizen observatories (four European and two African) and (b) and the demonstration of technological feasibility, sustainable implementation, and societal and economic value of such citizen observatories.

Co-funded by the European Union, Switzerland and the United Kingdom

LANDSENSE is an EU Horizon 2020 Citizen Observatory project that ran between September 2016 and December 2020 (LANDSENSE, 2020). The project aim was to empower and enable communities to monitor their environments, specifically focusing on land cover and land use. LANDSENSE lead a number of crowdsourcing and community-sourcing campaigns to collect new and validate existing datasets, including perceptions of urban green space quality in Amsterdam, updated authoritative LULC data in Toulouse and surrounding area, land cover/land use in Vojvodina (Serbia), interpretations of cropland field around the world, and more.

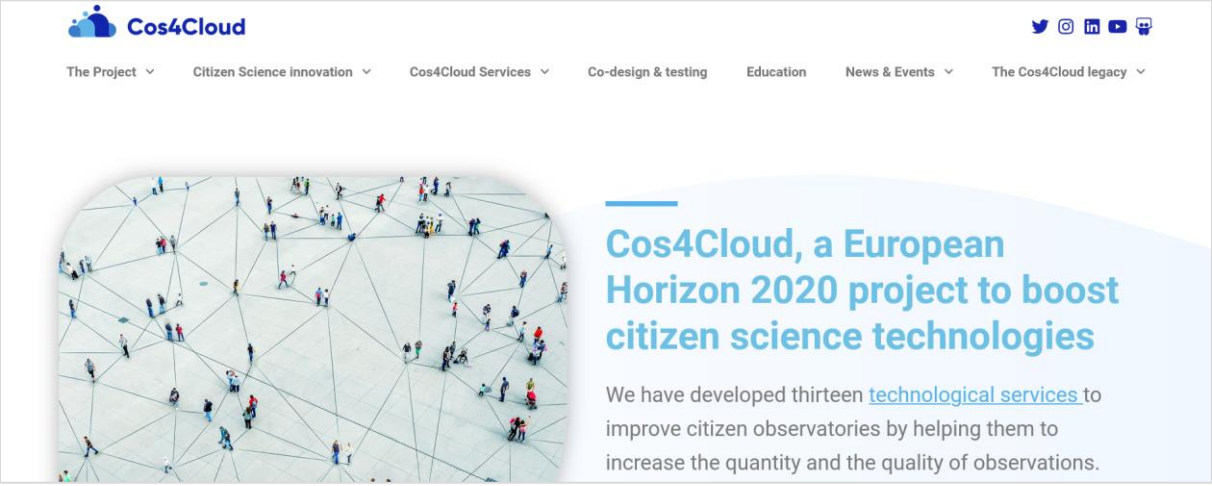


Figure 61: Cos4Clouds project home page.

Cos4Cloud is an EU Horizon 2020 project that ran between November 2019 and February 2023 (Cos4Cloud, 2022b). The key goal of the project was to co-design and develop tools to address the challenges of maintaining citizen observatories and the quality of citizen science data. The project has developed thirteen services, some of which are highlighted here (Cos4Cloud, 2022c), including the ***STAplus*** best practice document that culminated in an international standard after the project ended as described earlier.

Cos4Bio and ***Cos4Env*** are services for integration of biodiversity and environmental data respectively from multiple citizen observatories. These services allow to search and download datasets that combine data from different citizen observatories.

MOBIS is a service that allows creating citizen science apps (Android or iOS) for reporting environmental and biodiversity observations. It provides a user-friendly interface and allows to collect and combine diverse data including photographs and low-cost sensor data linked to a mobile website or a native app platform.

Pl@ntNet-API is an API that facilitates integration of automatic plant identification features into mobile apps. The API uses Pl@ntNet database to ensure regular updates with new flora images and continuously improve visual recognition.

The ***MECODA*** (ModulE for Citizen Observatory Data Analysis) open collaborative repository gathers data analysis and visualisation tools specifically designed for citizen science data. The toolkit is based on Orange Data Mining (data visualisation, machine learning and data mining toolkit) and facilitates graphs and charts, map-based analysis, data-quality analysis, and more. The MECODA toolkit enables users from citizen observatories to apply exploratory visual data analysis without the need for expert data visualisation or data analysis knowledge. It also facilitates the creation of reproducible visual dataflows that can be shared and reused.

Co-funded by the European Union, Switzerland and the United Kingdom

Figure 62: CitiObs project home page.

CitiObs is a tree-year EU Horizon 2020 project that started in January 2023 (CitiObs, 2023). The project will work towards:

- enhancing existing and new citizen observatories to engage citizens and marginalised communities,
- adding value to environmental observations in the urban context,
- increasing and validating citizen observations of the urban environment as part of the existing in-situ Earth Observation systems,
- co-creating inclusive local actions for sustainability, and
- ensuring that Citizen Observatories data contributes to research and policy development towards the objectives of the European Green Deal.

The work on the STAplus best practice document has evolved into an international standard during the first months of this project (OGC, 2023b).

7 CONCLUSIONS

The successful development of the EU Green Deal Data Space will depend on the availability of FAIR data sources, including FAIR citizen science data. The importance of FAIR data principles is increasingly acknowledged within the field of citizen science; this can be demonstrated by EU-Citizen.Science promoting FAIR data principles in their information resources and training courses and recent research into citizen science data FAIRification (Alvarez *et al.*, 2022). More efforts are however needed to support citizen science projects in delivering FAIR data. In this section, we discuss existing gaps and potential opportunities for citizen science data to become more FAIR.

7.1 FAIR DATA SOURCES FOR GDDS

GBIF, which includes citizen science observations data, is actively working towards observing FAIR data principles and only accepts data contributions that align with FAIR. While GBIF is a potential biodiversity data source for the EU 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 provider but, for example, museums may contribute both official and citizen science observations.

Established citizen science initiatives like OSM and Sensor.Community do not directly strive to comply with FAIR data principles, but they closely align with FAIR by observing good practices of Open Data. Successful applications of OSM data covering all GDDS themes (climate change, circular economy, pollution, biodiversity and deforestation) already exist. Some examples include support for global climate resilience (HOT, 2023), studies of open heat islands (Dimitrov *et al.*, 2021), classification of local climate zones (Fonte *et*

Co-funded by the European Union, Switzerland and the United Kingdom

al., 2019), OSM CircularEconomy project (OSM, 2022b), environmental assessment studies (Kloog *et al.*, 2018), research into habitat fragmentation and disturbances (Bista *et al.*, 2021, Snell *et al.*, 2020), crowdsourcing mapathon for deforestation detection (Bratic and Brovelli, 2022), and urban forest mapping software (PlanIT GEO, 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 to support the GDDS users.

Sensor.Community data holds significance in the context of the GDDS focus on climate change and pollution, and can complement official environmental data sources at a local scale. To accomplish successful inclusion within the GDDS, Sensor.Community data will need to be semantically enriched using controlled vocabularies to ensure seamless integration with other sources. Additional APIs or service layers will also be required to 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).

As discussed in Section 3, data FAIRness in citizen science can be achieved using either top-down or bottom-up approach. GBIF follows a top-down approach by specifying rigorous standards for data contributions (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 EML and DwC-A (**I**), requiring creative common data licences and recording data provenance (**R**) (GBIF, 2022).

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

7.2 INDEPENDENT CITIZEN SCIENCE PROJECTS

Citizen science projects that operate independently from larger initiatives may lack awareness of FAIR guiding principles, struggle to select suitable standards and tools, or fail to recognise the value of sharing their data outside of the project. In a commercial environment, there may be a full suite of useful tools and applications to support end-to-end data lifecycle, however, they may become FAIR only in their closed data ecosystem (e.g. ArcGIS). If data owners wish to transfer, integrate or share their data with external sources, licensing can introduce limitations. On the other hand, using open source platforms and tools that support citizen science projects (e.g., ODK) is limited with options required for data lifecycle management in a FAIR way. Therefore, each project needs to select and combine different tools by purpose from different providers, resulting in more challenges to achieve a seamless flow of FAIR data.

A vast number of domain-specific controlled vocabularies and other semantic resources exist and are used by the scientific community and by some established citizen science initiatives (e.g., Darwin Core by GBIF). Independent citizen science projects, however, rarely apply standardised semantic tagging because they are either unaware of its importance or unsure which resources to choose. This presents a major gap for data discovery and interoperability, especially cross domain. Tools for selecting and extending semantic resources, similar to the EcoPortal Annotator/Mapper/Recommender tools for the ecological domain, will need to be developed for a wide range of domains to support the citizen science community. While certain semantic resources may be too complex for citizen science initiatives, tools for building custom vocabularies or ontologies can be used to create simplified versions while preserving links to the original sources of terms (e.g., Semantic Treehouse vocabulary hub (van den Berg, 2023)).

Co-funded by the European Union, Switzerland and the United Kingdom

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 the project funding terminates. Open repositories like Zenodo and Pangea can be used to publish and share citizen science datasets, but search capabilities are limited and it is not possible to search for citizen science data specifically. Another limitation is that these 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 transparent 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.

There is also a potential for citizen science data to be integrated in environmental Research Infrastructures or e-infrastructures that can serve as the intermediaries connecting to the European GDDS. In exchange for citizen science data, such infrastructures shall increase technical and semantic services to support citizen science projects in meeting high Technological Readiness Levels in operational environments. Future citizen science project calls should include a strategic plan on how long-term sustainability will be ensured for services developed during the project period, by connecting them with specific environmental Research Infrastructures such as LifeWatch ERIC, eLTER and many others illustrated in the environmental cluster of RIs (Figure 63 ESFRI Roadmap 2021). In the past, COS4Cloud made a similar effort to bring Citizen Science to EOSC.

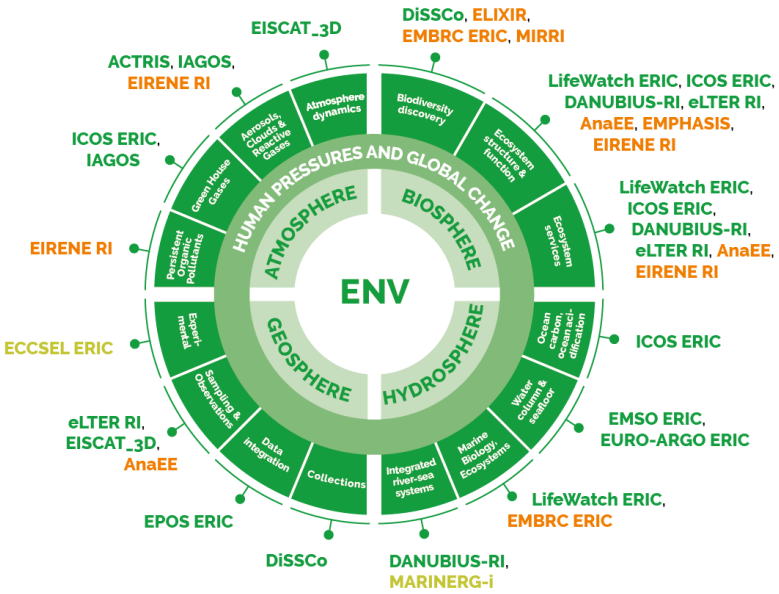


Figure 63: ESFRI Roadmap 2021 – Environmental Research Infrastructures (Source: ESFRI 2021).

The findings outlined in this document will be investigated further to identify any new emerging platforms, tools, standards, and semantic resources relevant to citizen science. Any new additional findings will be published in the Deliverable 2.4 of this project.

To investigate what standards and tools citizen science projects use in practice, this project will conduct a structured online survey that will be distributed via major EU citizen science platforms and national networks. The results of this survey will help to understand what gaps are perceived by citizen science initiatives and what tools and resources will need to be developed to support data FAIRness and data integration into the GDDS.

Ongoing project activities will also investigate the gap in federating citizen science projects and in producing catalogues of citizen science datasets. This also includes the question of whether any citizen science datasets should be contributed to the GDDS, or whether the GDDS should focus on high value datasets as currently identified by the European Commission (GREAT, 2023).

Co-funded by the European Union, Switzerland and the United Kingdom

8 ANNEX I: CITIZEN SCIENCE PROJECTS DISCOVERY PLATFORMS

Table 5: Citizen science projects discovery platforms.

Citizen Science Projects Discovery Platforms				
Platform / Description	**Governing Body / Funding**	**Domain / Coverage**	**Functions**	**Notes**
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.
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.

Co-funded by the European Union, Switzerland and the United Kingdom

9 ANNEX II: NATIONAL EUROPEAN CITIZEN SCIENCE NETWORKS

Table 6: National European citizen science networks.

National Citizen Science Networks				
Platform / Description	**Governing Body / Funding**	**Domain / Coverage**	**Functions**	**Notes**
Vera (EU) https://vera.operas-eu.org/ European network that aims to facilitate collaboration and promote CitSci research. Lists 14 projects.	Developed by COESO (https://coeso.hypotheses.org/). Funded by the EU Horizon 2020 programme.	Social Sciences and Humanities / Europe	**Main:** EU-wide network **Additional:** Project discovery Project hosting Collaborator discovery Funding opportunities	Project discovery and hosting is aimed at facilitating collaboration and networking, not project participant recruitment.
Österreich forscht https://www.citizen-science.at/en/ Austrian citizen science network. Founded in 2017. Lists 72 active projects.	Supported by the University of Natural Resources and Life Sciences, Vienna.	Any/ Austria	**Main:** National network **Additional:** Project discovery CitSci quality criteria	Has a well-defiled transparent quality criterion for hosted projects: https://zenodo.org/record/1161953
Norwegian Network for Citizen Science Forskningsradet https://www.forskningsradet.no/forskningspolitikk-strategi/nasjonalt-nettverk-folkeforskning/ The Norwegian Network for Citizen Science. Aims to promote CitSci in Norway, coordinate activities, and help exchange experience.	Coordinated by Norwegian Research Council.	Any / Norway	**Main:** National network **Additional:** Under development	Website is only available in Norwegian language. Currently under development.
Citizen Science Nederland https://www.cs-nl.network/ Launched in 2022 by the National Program Open Science.	Awaiting funding from the Open Science Regieorgaan in the Netherlands.	Any / Netherlands	**Main:** National network **Additional:**	In very early development.

Co-funded by the European Union, Switzerland and the United Kingdom

			Under development	
Scivil (Citizen Science Vlaanderen) https://www.scivil.be/ Works towards promoting CitSci and connecting stakeholders in Flanders and Belgium.	Scivil is housed in Imec's offices in Leuven: https://www.imec.be/nl. Advised by a committee of 22 citizen science experts and representatives of Flemish research institutions.	Any/ Flanders & Belgium	**Main:** National network **Additional:** Project discovery Collaborator discovery Funding opportunities Information resources	Project discovery platform supports both Dutch and English languages. Number of registered projects is not specified. Engaged in Air Quality projects.
Iedereen Wetenschapper https://www.iedereenwetenschapper.be/ CitSci network in Flanders and the Netherlands.	Funding sources are unknown.	Any / Flanders & Netherlands	**Main:** National network **Additional:** Project discovery	Works in close collaboration with Citizen Science Vlaanderen. Website is only available in Dutch language. Number of registered projects is not specified.
Citizen Science Netværket https://www.citizenscience.dk/ Denmark-wide CitSci network that focuses on promoting CitSci projects and facilitating networking and collaboration.	Funding sources are unknown.	Any / Denmark	**Main:** National network **Additional:** Under development	In very early stages of development.
Citizen Science Italy https://www.citizenscience.it/ CitSci network for coordinating initiatives in Italy. Aims to achieve formal recognition of CitSci activities at the national level.	Awaiting potential funding from the National Biodiversity Future Center.	Any / Italy	**Main:** National network **Additional:** Under development	In very early stages of development.
Rede Portuguesa de Ciência Cidadã https://www.cienciacidada.pt/ Portuguese CitSci network. Lists 3 CitSci projects.	Run by volunteers and does not yet have sources of funding.	Any/ Portugal	**Main:** National network **Additional:** Project discovery	Website is only available in Portuguese.
CS-Center for nature, sustainability and digitalization	An 18-month project funded by Deutsche Bundesstiftung Umwelt	Nature Conservation &	**Main:** CitSci Centre	Focuses on digitalisation challenges and aims to increase digital and technological competences.

Co-funded by the European Union, Switzerland and the United Kingdom

https://www.museumfuernaturkunde.berlin/en/science/development-a-citizen-science-center-nature-sustainability-and-digitalization Project aims to establish a CitSci centre in Germany that will support nature conservation and sustainability CitSci projects.	and coordinated by Musem of Natural History Berlin.	Sustainability / Germany	**Additional:** Under development	
OeAD Center for Citizen Science https://zentrumfuercitizenscience.at/en/ Information, advisory and service centre for CitSci.	OeAD – Austria's Agency for Education and Internationalisation	Any / Austria	**Main:** Service centre **Additional:** Information resources Funding opportunities Collaborator discovery	Citizen Science Contact Persons network: https://zentrumfuercitizenscience.at/en/networking/contacts-at-research-institutions
Medborgar Forskning.Se https://medborgarforskning.se/ Swedish CitSci network that offers tool and guidelines to CitSci researchers and other stakeholders.	The networks' website was launched in 2021 by the ARenas for Cooperation through citizen Science (ARCS) project that received funding from Vinnova (Sweden's innovation agency), the University of Gothenburg, the Swedish University of Agricultural Sciences, and Umeå University.	Any / Sweden	**Main:** National network **Additional:** Project discovery Information resources Forum	Medborgar Forskning.se uses EU-Citizen.Science platform as a project discovery portal for projects located in Sweden. Website is in Swedish language, except for home page that is also available in English.
CitizenScience.cz https://www.citizenscience.cz/ Czech Republic CitSci network.	Coordinated by the Czech Academy of Sciences and Tomas Bata University in Zlin.	Any / Czech Republic	**Main:** National network **Additional:** Project discovery	Website is only available in Czech language. Number of registered projects is not specified. Project discovery platform only supports search by keywords.

Co-funded by the European Union, Switzerland and the United Kingdom

10 ANNEX III: CITIZEN SCIENCE INFORMATION RESOURCES AND TRAINING MATERIALS

Table 7: Citizen science information resources and training materials.

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.
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	The course highlights such topics as data sharing with the scientific community and with the project participants.

Co-funded by the European Union, Switzerland and the United Kingdom

		Data sharing	
CitieS-Health Toolkit https://citizensciencetoolkit.eu/about/ A collection of resources and tools to support individual citizens, communities, scientists, public authorities, and other organisations in defining, designing, deploying, and disseminating participatory studies that tackle environmental and health issues.	Environment Health / Project leaders Individual citizens Communities Scientists Public authorities Other organisations	CitieS-Health / Various resources covering: • Identification • Co-design • Deployment • Action	The toolkit resources are organised around four main stages of participatory studies: Identification, Co-design, Deployment, and Action. The toolkit contains 31 resources. New resources can be submitted and published on the platform under the creative common licence CC BY-SA 4.0, subject to review and approval.
CitizenScience.gov Toolkit https://www.citizenscience.gov/toolkit/howto/ A collection of informational resources for crowdsourcing and citizen science projects.	Any / Project leaders	CitizenScience.gov / Project scoping Project design Community building Data management plan Project sustainability Project hosting Project improvement	The resources are grouped under five basic process steps: project scoping, design, building a community, managing data, and sustaining and improving project over time. Data management resource highlights Open Data Policy for data preservation and sharing: https://digital.gov/resources/open-data-policy-m-13-13/.
EPA Quality Assurance Toolkit https://www.epa.gov/participatory-science/quality-assurance-handbook-and-toolkit-participatory-science-projects A collection of resources that defines best practices in documenting the quality of environmental data.	Environment / Citizen science organisations Project leaders	CitizenScience.gov / Data quality Data quality documentation Data quality assurance templates	The toolkit outlines quality assurance requirements, and offers quality assurance templates (e.g., data quality objectives and indicators, volunteer training).
Data Quality Resource Compendium https://citizenscience.org/data-resources/ A living collection of data quality guidance documents, manuals, and workbooks designed for citizen science.	Any / Project leaders	CitizenScience.org / Data quality assurance Project planning Data management Data collection Data documentation Data preservation Data integration Data analysis	The resources in the collection focus on data quality assurance at all stages of citizen science project lifecycle. The resource is maintained by the CSA Data and Metadata Working Group.

Co-funded by the European Union, Switzerland and the United Kingdom

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 / 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.

11 ANNEX IV: TOOLS THAT CAN SUPPORT CITIZEN SCIENCE PROJECT LIFECYCLE

Table 8: Tools that can support citizen science project lifecycle.

Tools that can support citizen science project lifecycle				
Tool / Description	**Governing Body / Funding**	**Domain**	**Functions**	**Notes**
CitSci 2.0 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.
Anecdata https://anecdata.org Free online citizen science project hosting platform	Founded by the Community Lab at the MDI Biological Laboratory in Bar Harbor, Maine.	Any	**Main:** Project hosting **Additional:** Project discovery (limited)	It is well suited for more complex biodiversity protocols such as recording absence data, water quality monitoring, litter recording and clean up, and

Co-funded by the European Union, Switzerland and the United Kingdom

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.			Custom data sheets Data collection Mobile app (Android, iOS) Participation management Data sharing Access management Data visualisation on map	collection of non-biodiversity image observations. Integrates with SciStarter Project and observation data search are limited. Not compliant to FAIR.
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.
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.
Open Data Kit (ODK) https://getodk.org/ Open source data collection toolkit for mobile devices.	Managed by ODK Inc. Funded via ODK Cloud subscription plans.	Any	**Main:** Data collection **Additional:** Data storage	Support efficient and reliable data collection in offline or low-connectivity. Additionally offers ODK Cloud – a fully managed subscription-based platform.

Co-funded by the European Union, Switzerland and the United Kingdom

Epicollect5 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 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.
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.
SPOTTERON https://www.spotteron.net/ commercial web-based platform that facilitates citizen science projects and environmental monitoring initiatives	Owned by a commercial company SPOTTERON.	Any	**Main:** Data collection **Additional:** Project creation Project hosting Custom datasheets Mobile app (Android, iOS) Data sharing Data quality Customisable project website Community support	Commercial product.
ArcGIS Open Data https://www.esri.com/en-us/arcgis/products/arcgis-open-data	Managed by Esri commercial company.	Any	**Main:** Customisable project website **Additional:**	Commercial product. Needs to be purchased as part of ArcGIS Online toolkit.

Co-funded by the European Union, Switzerland and the United Kingdom

Facilitates creation of custom sites using existing ArcGIS templates			Data storage Advance data analysis Interactive maps creation Multiuser collaboration	
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.
ArcGIS Conservation Outreach https://doc.arcgis.com/en/arcgis-solutions/11.0/reference/introduction-to-conservation-outreach.htm Designed for management of protected areas to monitor wildlife conflicts	Managed by Esri commercial company.	Biodiversity	**Main:** Management of protected areas **Additional:** Incident reports Monitoring of conflict incidents Management of conflict resolutions	Commercial product. Needs to be purchased as part of ArcGIS Online toolkit. Requires ArcGIS Online and ArcGIS Survey123 to be deployed.

Co-funded by the European Union, Switzerland and the United Kingdom

ArcGIS StoryMaps https://doc.arcgis.com/en/arcgis-solutions/latest/reference/introduction-to-conservation-outreach.htm Map-based data visualisation product that allows to create influential interactive maps	Managed by Esri commercial company.	Any	**Main:** Map-based visualisations	Commercial product. Needs to be purchased as part of ArcGIS Online toolkit.
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	
EnvThes (Environmental Thesaurus) https://vocabs.lter-europe.net/envthes/en/ A controlled vocabulary of semantically well-defined terms that are related to long-term ecological research, monitoring and experiments.	Maintained by LTER-Europe	Ecological research	**Main:** Controlled vocabulary	Based on NASA Sweet Units, Catalogue of Life, EUNIS Habitats, INSPIRE spatial data themes.
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
EcoPortal https://ecoportal.lifewatch.eu/ The LifeWatch ERIC repository of semantic resources for ecology and related domains.	Managed by LifeWatch ERIC.	Ecology and related domains	**Main:** Controlled vocabulary	Provides additional tools: EcoPortal Annotator EcoPortal Mappings EcoPortal Recommender

Co-funded by the European Union, Switzerland and the United Kingdom

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.
PANGEA https://www.pangaea.de/ An Open Access library and a publishing platform for georeferenced data. The platform Established in the 1990s.	Hosted by the Alfred Wegener Institute. Supported by The European Commission, Federal Ministry of Education and Research (BMBF), Deutsche Forschungsgemeinschaft (DFG) and International Ocean Discovery Program (IODP).	Earth and life sciences	**Main:** Data repository	Review of submissions can take up to several months depending on complexity.
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-sensing-entities An international standard for interconnecting heterogeneous Internet of Things (IoT) devices. The first version published in 2016.	Maintained by the Open Geospatial Consortium.	Any	**Main:** Data standard	
OGC STAplus 1.0 https://docs.ogc.org/DRAFTS/22-022.html	Maintained by the Open Geospatial Consortium.	Any	**Main:** Data standard	Specific to citizen science.

Co-funded by the European Union, Switzerland and the United Kingdom

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.				
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.

Co-funded by the European Union, Switzerland and the United Kingdom

12 REFERENCES

AD4GD Project 101061001. GRANT AGREEMENT. European Research Executive Agency (REA)

ALA. 2023. "The Atlas of Living Australia home page." Accessed August 30, 2023. https://www.ala.org.au.

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.

Anecdata. 2023. "Home page." Accessed August 30, 2023. https://anecdata.org.

ArcGIS. 2023a. "ArcGIS Online." Accessed August 30, 2023. https://www.arcgis.com/index.html.

ArcGIS. 2023b. "ArcGIS Open Data." Accessed August 30, 2023. https://www.esri.com/en-us/arcgis/products/arcgis-open-data.

ArcGIS. 2023c. "ArcGIS Survey 123." Accessed August 30, 2023. https://survey123.arcgis.com/.

ArcGIS. 2023d. "ArcGIS QuickCapture." Accessed August 30, 2023. https://www.esri.com/en-us/arcgis/products/arcgis-quickcapture/overview.

ArcGIS. 2023e. "Introduction to Community Science." Accessed August 30, 2023. https://doc.arcgis.com/en/arcgis-solutions/latest/reference/introduction-to-community-science.htm.

ArcGIS. 2023f. "ArcGIS StoryMaps." Accessed August 30, 2023. https://doc.arcgis.com/en/arcgis-solutions/latest/reference/introduction-to-conservation-outreach.htm.

ArcGIS. 2023g. "Introduction to Conservation Outreach." Accessed August 30, 2023. https://www.esri.com/en-us/arcgis/products/arcgis-storymaps/overview.

B-Cubed. 2023. "Partners." Accessed August 30, 2023. https://b-cubed.eu/partners.

BGRM. 2023. "France's reference public institution for Earth Science applications." Accessed August 30, 2023. https://www.brgm.fr/en.

Bista, Damber, Greg S. Baxter, Nicholas J. Hudson, Sonam Tashi Lama, and Peter John Murray. "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.

BODC. 2018. "The BODC Parameter Usage Vocabulary (PUV) semantic model exposed." Accessed August 30, 2023. https://www.bodc.ac.uk/resources/vocabularies/parameter_codes/documents/BODC_P01_PUV_semantic_model_Aug19.pdf

BSCS. 2023a. "Invitations to Inquiry with Data using FieldScope." Accessed August 30, 2023. https://bscs.org/resources/invitations-to-inquiry/.

Co-funded by the European Union, Switzerland and the United Kingdom

BSCS. 2023b. "BSCS Home page." Accessed August 30, 2023. https://bscs.org/.

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 43: 231-238.

Chapman Arthur D. 2020. "Current Best Practices for Generalizing Sensitive Species Occurrence Data." Accessed August 30, 2023. https://docs.gbif.org/sensitive-species-best-practices/master/en/.

CERN. 2023. "CERN Analysis Preservation: Capture, preserve and reuse physics analyses." Accessed August 30, 2023. https://analysispreservation.cern.ch.

CienciaCidada. 2023. "Home page." Accessed August 30, 2023. https://www.cienciacidada.pt/.

CitieS-Health. 2023. "About this toolkit." Accessed August 30, 2023. https://citizensciencetoolkit.eu/about/.

CitiObs. 2023. "Home page." Accessed August 30, 2023. https://citiobs.eu/.

CitSci. 2023. "Helping you do great science: Create projects, build datasheets, collect data, and view results in real-time." Accessed August 30, 2023. https://www.citsci.org/.

Citizen Science Zürich. 2023. "Citizen Science Project Builder." Accessed August 30, 2023. https://lab.citizenscience.ch/en/tools/projectbuilder.

CitizenScience.gov. 2023. "CitizenScience.gov Home page." Accessed August 30, 2023. https://www.citizenscience.gov.

Citizen Science Italy. 2023. "Citizen Science Italy." Accessed August 30, 2023. https://www.citizenscience.it/.

Citizen Science Netværket. 2021. "Citizen Science Netværket." Accessed August 30, 2023. https://www.citizenscience.dk/.

COESO. 2023. "COESO project." Accessed August 30, 2023. https://coeso.hypotheses.org/.

Cos4Bio. 2022. "Cos4Bio home page." Accessed August 30, 2023. https://cos4cloud-eosc.eu/services/cos4bio/.

Cos4Cloud. 2022a. "Cos4Cloud STAplus Viewer App." Accessed August 30, 2023. https://cos4cloud.demo.secure-dimensions.de/combi-viewer-app/.

Cos4Cloud. 2022b. "Home page." Accessed August 30, 2023. https://cos4cloud-eosc.eu/.

Cos4Cloud. 2022c. "Cos4Cloud Services." Accessed August 30, 2023. https://cos4cloud-eosc.eu/services/.

Cos4Env. 2022. "Cos4Env home page." Accessed August 30, 2023. https://cos4cloud-eosc.eu/cos4env/.

Co-funded by the European Union, Switzerland and the United Kingdom

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.

CS-NL. 2023. "WELKOM." Accessed August 30, 2023. https://www.cs-nl.network/.

CSA. 2021a. "PPSR Core: A Data Standard for Public Participation in Scientific Research (Citizen Science)." Accessed August 30, 2023. https://core.citizenscience.org/.

CSA. 2021b. "citizen-science-association / ppsr-core." Accessed August 30, 2023. https://github.com/citizen-science-association/ppsr-core/tree/master.

DataCove. 2019. "Integration of O&M data in the INSPIRE SDI - Benefits, challenges and prospects." Accessed August 30, 2023. http://www.datacove.eu/inspire/o-m-in-inspire-sdi/.

DEFRA. 2023. "DEFRA Air Pollution Glossary." Accessed August 30, 2023. https://uk-air.defra.gov.uk/air-pollution/glossary.

DEIMS. 2023. "Welcome to DEIMS-SDR." Accessed August 30, 2023. https://deims.org/.

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.

eBird. 2023. "Home page." Accessed August 30, 2023. https://ebird.org/home.

EC. 2015. "INSPIRE registry: Environmental monitoring facilities." Accessed August 30, 2023. https://inspire.ec.europa.eu/theme/ef.

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. 2023a. "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.

EC. 2023b. "European Green Deal." Accessed August 30, 2023. https://www.consilium.europa.eu/en/policies/green-deal/.

EC. 2023c. "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.

Co-funded by the European Union, Switzerland and the United Kingdom

EC. 2023d. "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. 2023e. "OGC SensorThings API as an INSPIRE download service." Accessed August 30, 2023. https://inspire.ec.europa.eu/good-practice/ogc-sensorthings-api-inspire-download-service.

EcoPortal. 2023a. "Environmental Thesaurus." Accessed August 30, 2023. https://ecoportal.lifewatch.eu/ontologies/ENVTHES

EcoPortal. 2023b. "Welcome to EcoPortal, the LifeWatch ERIC repository of semantic resources for ecology and related domains." Accessed August 30, 2023. https://ecoportal.lifewatch.eu/.

EcoPortal. 2023c. "EcoPortal API Documentation." Accessed August 30, 2023. http://ecoportal.lifewatch.eu:8080/documentation.

ECSA. 2015. "Ten Principles of Citizen Science". Berlin. Accessed August 30, 2023. http://doi.org/10.17605/OSF.IO/XPR2N

Eionet. 2023. "Eionet Data Dictionary." Accessed August 30, 2023. https://dd.eionet.europa.eu/.

eLTER. 2023a. "EnvThes - Thesaurus for long term ecological research, monitoring and experiments." Accessed August 30, 2023. https://vocabs.lter-europe.net/envthes/en/.

eLTER. 2023b. "eLTER_CL is a thesaurus for controlled lists used by the eLTER community." Accessed August 30, 2023. https://vocabs.lter-europe.net/elter_cl/en/.

eLTER. 2023c. "Standard observations (coming soon)." Accessed August 30, 2023. https://vocabs.lter-europe.net/so/en/.

Epicollect5. 2023. "Home page." Accessed August 30, 2023. https://five.epicollect.net/.

ESFRI. 2021. "ESFRI Roadmap 2021 - Strategy Report on Research Infrastructures, Part 2 Landscape Analysis - Section 1 Environment." Accessed August 30, 2023. https://roadmap2021.esfri.eu/media/1259/rm21-part-2-env.pdf.

EU-Citizen.Science. 2023a. "Home page." Accessed August 30, 2023. https://eu-citizen.science/.

EU-Citizen.Science. 2023b. "Swagger API." Accessed August 30, 2023. https://eu-citizen.science/swagger/.

Farrell, Eimear, Marco Minghini, Alexander Kotsev, Josep Soler Garrido, Brooke Tapsall, Marina Micheli, Monica Posada Sanchez, Serena Signorelli, Alessio Tartaro, Jaime Bernal Cereceda, Michele Vespe, Margherita Di Leo, Bruno Carballa Smichowski, Robin Smith, Sven Schade, Katarzyna Pogorzelska, Lorenzo Gabrielli and Davide de Marchi. 2023. "European Data Spaces - Scientific Insights into Data Sharing and Utilisation at Scale." Joint Research Centre (Seville site). doi:10.2760/301609.

FieldScope. 2023. "Home page." Accessed August 30, 2023. https://www.fieldscope.org/.

Co-funded by the European Union, Switzerland and the United Kingdom

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.

Forskningsradet. 2023. "Nasjonalt nettverk for folkeforskning." Accessed August 30, 2023. https://www.forskningsradet.no/forskningspolitikk-strategi/nasjonalt-nettverk-folkeforskning/.

Fraunhofer. 2023. "FROST-Server - An open source implementation of OGC SensorThings API." Accessed August 30, 2023. https://www.iosb.fraunhofer.de/en/projects-and-products/frost-server.html.

GBIF. 2023a. "GBIF Global Biodiversity Information Facility: Free and open access to biodiversity data." Accessed August 30, 2023. https://www.gbif.org/.

GBIF. 2023b. "GBIF Registered Extensions." Accessed August 30, 2023. https://rs.gbif.org/extensions.

GBIF. 2023c. "Diversifying the GBIF Data Model." Accessed August 30, 2023. https://www.gbif-uat.org/composition/HjlTr705BctcnaZkcjRJq/gbif-new-data-model.

GBIF. 2023d. "The Darwin Core Archive Assistant." Accessed August 30, 2023. http://tools.gbif.org/dwca-assistant/.

GBIF. 2023e. "The Integrated Publishing Toolkit (IPT)." Accessed August 30, 2023. https://www.gbif-uat.org/ipt.

GBIF. 2023f. "The GBIF Validator Tool." Accessed August 30, 2023. https://www.gbif-uat.org/tools/data-validator.

GBIF. 2023g. "BioCollect." Accessed August 30, 2023. https://www.gbif.org/publisher/1d9d1d61-7871-44b4-9ceb-0bc2fe809d2e.

GBIF. 2023h. "iNaturalist Research-grade Observations." Accessed August 30, 2023. https://www.gbif.org/dataset/50c9509d-22c7-4a22-a47d-8c48425ef4a7.

GBIF. 2023i. "gbif-observations-dwca." Accessed August 30, 2023. http://www.inaturalist.org/observations/gbif-observations-dwca.zip.

GBIF. 2023j. "EOD – eBird Observation Dataset." Accessed August 30, 2023. https://www.gbif.org/dataset/4fa7b334-ce0d-4e88-aaae-2e0c138d049e.

GCUBE. 2013. "Darwin Core Terms." Accessed August 30, 2023. https://gcube.wiki.gcube-system.org/gcube/Darwin_Core_Terms/.

Geo-Wiki. 2023. "Home paeg." Accessed August 30, 2023. https://www.geo-wiki.org.

Geofabrik. 2023. "OpenStreetMap Data Extracts." Accessed August 30, 2023. https://download.geofabrik.de/.

Co-funded by the European Union, Switzerland and the United Kingdom

GO FAIR. 2016. "FAIR Principles." Accessed August 30, 2023. https://www.go-fair.org/fair-principles/.

GLOBE. 2023a. "The Global Learning and Observations to Benefit the Environment home page." Accessed August 30, 2023. https://www.globe.gov.

GLOBE. 2023b. "Observer." Accessed August 30, 2023. https://observer.globe.gov.

GLOBE. 2023c. "Earth as a System Community." Accessed August 30, 2023. https://www.globe.gov/web/earth-systems/community.

GLOBE. 2023d. "Protocol eTraining." Accessed August 30, 2023. https://www.globe.gov/get-trained/protocol-etraining

GLOBE. 2023e. "Required eTraining Modules by Sphere/Protocol." Accessed August 30, 2023. https://www.globe.gov/get-trained/protocol-etraining/protocol-requirements.

GLOBE. 2023f. "Visualization System." Accessed August 30, 2023. https://www.globe.gov/globe-data/visualize-and-retrieve-data.

GLOBE. 2023g. "GLOBE Advanced Data Access Tool." Accessed August 30, 2023. https://www.globe.gov/globe-data/retrieve-data.

GLOBE. 2023h. "GLOBE API." Accessed August 30, 2023. https://www.globe.gov/globe-data/globe-api.

GLOBE. 2023i. "Get Data." Accessed August 30, 2023. https://observer.globe.gov/get-data#data.

GREAT. 2023. Home page." Accessed August 30, 2023. https://www.greatproject.eu/.

GroundTruth 2.0. 2019. "Home page." Accessed August 30, 2023. https://gt20.eu/.

GROW. 2019. "Home page." Accessed August 30, 2023. https://growobservatory.org/.

HackAIR. 2023. "Home page." Accessed August 30, 2023. https://platform.hackair.eu/.

HeiGit. 2023. "OSM Landuse / Landcover ." Accessed August 30, 2023. http://osmlanduse.org/.

HOT, 2023. "How HOT is Supporting Climate Resilience Around the World." Accessed August 30, 2023. https://www.hotosm.org/updates/how-hot-is-supporting-climate-resilience-around-the-world/.

Ideas for Change. 2023. "Home page." Accessed August 30, 2023. https://www.ideasforchange.com/.

IedereenWetenschapper. 2023. "Home page." Accessed August 30, 2023. https://www.iedereenwetenschapper.be/.

iNaturalist. 2023. "Home page." Accessed August 30, 2023. https://www.inaturalist.org.

Co-funded by the European Union, Switzerland and the United Kingdom

iNaturalist. 2023b. "Developers." Accessed August 30, 2023. https://www.inaturalist.org/pages/developers.

iNaturalist. 2023c. "iNaturalist Network." Accessed August 30, 2023. https://www.inaturalist.org/sites/network.

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.

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.

KSA. 2023. "Globe Protocols." Accessed August 30, 2023. https://ksa.go.ke/pages/globe-protocols.

LANDSENSE. 2020. "Home page." Accessed August 30, 2023. https://landsense.eu/.

Liu, Hai-Ying, Daniel Dörler, Florian Heigl, and Sonja Grossberndt. "Citizen science platforms." The Science of Citizen Science 22 (2021): 439-459.

Medborgarforskning. 2023. "Vad är medborgarforskning?" Accessed August 30, 2023. https://medborgarforskning.se/.

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.

Museum Fuer Naturkunde. 2023. "Development of a Citizen Science Center for Nature, Sustainability and Digitalization." Accessed August 30, 2023. https://www.museumfuernaturkunde.berlin/en/science/development-a-citizen-science-center-nature-sustainability-and-digitalization.

Natusfera. 2023. "Natusfera home page." Accessed August 30, 2023. https://natusfera.gbif.es/?locale=en/.

NERC. 2023a. "The NERC Vocabulary Server (NVS)." Accessed August 30, 2023. https://vocab.nerc.ac.uk/.

NERC. 2023b. "The NERC Vocabulary Server (NVS): SPARQL." Accessed August 30, 2023. http://vocab.nerc.ac.uk/sparql/.

NERC. 2023c. "The NERC Vocabulary Server (NVS)." Accessed August 30, 2023. https://vocab.nerc.ac.uk/docs/.

Co-funded by the European Union, Switzerland and the United Kingdom

ODK. 2023. "Collect data anywhere." Accessed August 30, 2023. https://getodk.org/.

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.

OGC. 2023a. "OGC SensorThings API: Overview." Accessed August 30, 2023. https://www.ogc.org/standard/sensorthings/.

OGC. 2023b. "OGC SensorThings API Extension: STAplus 1.0." Accessed August 30, 2023. https://docs.ogc.org/is/22-022r1/22-022r1.html.

OpenAIRE. 2023. "Discover open linked research." Accessed August 30, 2023. https://explore.openaire.eu/.

OpenMapTiles. 2023. "penmaptiles / openmaptiles-tools." Accessed August 30, 2023. https://github.com/openmaptiles/openmaptiles-tools#multi-streamed-osm-data-downloader/.

OSM. 2020. "Overpass turbo / GeoJSON." Accessed August 30, 2023. https://wiki.openstreetmap.org/wiki/Overpass_turbo/GeoJSON.

OSM. 2022a. "Downloading data." Accessed August 30, 2023. https://wiki.openstreetmap.org/wiki/Downloading_data.

OSM. 2022b. "WikiProject CircularEconomy." Accessed August 30, 2023. https://wiki.openstreetmap.org/wiki/WikiProject_CircularEconomy.

OSM. 2023a. "OpenStreetMap." Accessed August 30, 2023. https://www.openstreetmap.org.

OSM. 2023b. "Copyright and Licence." Accessed August 30, 2023. https://www.openstreetmap.org/copyright.

OSM. 2023c. "List of OSM-based services." Accessed August 30, 2023. https://wiki.openstreetmap.org/wiki/List_of_OSM-based_services.

OSM. 2023d. "Elements." Accessed August 30, 2023. https://wiki.openstreetmap.org/wiki/Elements.

OSM. 2023e. "Overpass API." Accessed August 30, 2023. https://wiki.openstreetmap.org/wiki/Overpass_API.

OSM. 2023f. "API v0.6." Accessed August 30, 2023. https://wiki.openstreetmap.org/wiki/API_v0.6.

Österreich forscht. 2023. "Welcome at Österreich forscht." Accessed August 30, 2023. https://www.citizen-science.at/en/.

PANGEA. 2023a. "Welcome to PANGAEA Data Publisher." Accessed August 30, 2023. https://www.pangaea.de.

Co-funded by the European Union, Switzerland and the United Kingdom

PANGEA. 2023b. "Data Model." Accessed August 30, 2023. https://wiki.pangaea.de/wiki/Data_model.

PANGEA. 2023c. "Tools." Accessed August 30, 2023. https://www.pangaea.de/tools/.

Peng, G. 2023. "Finding harmony in FAIRness." Eos. Accessed August 30, 2023. https://eos.org/opinions/finding-harmony-in-fairness.

Pl@ntNet. 2023. "Pl@ntNet home page." Accessed August 30, 2023. https://plantnet.org/en/.

Planet OSM. 2023a. "OpenStreetMap stats." Accessed August 30, 2023. https://planet.openstreetmap.org/statistics/data_stats.html.

Planet OSM. 2023b. "Home page." Accessed August 30, 2023. https://planet.openstreetmap.org/.

Planet OSM. 2023c. "Complete OSM Data History." Accessed August 30, 2023. https://planet.openstreetmap.org/planet/full-history/.

PlanIT GEO. 2023. "A Better Way to Map and Manage Your Urban Forest." Accessed August 30, 2023. https://marketing.planitgeo.com/treeplotter-software-uk.

Pybossa. 2023. "Home page." Accessed August 30, 2023. https://pybossa.com/support/.

SamenMeten. 2023. "Dataportaal." Accessed August 30, 2023. https://samenmeten.rivm.nl/dataportaal/.

Samen voor Zuivere Lucht. 2023. "Dataportaal." Accessed August 30, 2023. https://samenvoorzuiverelucht.eu/dataportaal/.

SCENT. 2019. "Home page." Accessed August 30, 2023. https://scent-project.eu/.

Schentz, Herbert, Johannes Peterseil, and Nic Bertrand. 2013. "EnvThes–interlinked thesaurus for long term ecological research, monitoring, and experiments." Proceedings of the 27th Conference on Environmental Informatics-Informatics for Environmental Protection, Sustainable Development and Risk Management. Shaker Verlag.

SciStarter. 2023a. "Home page." Accessed August 30, 2023. https://scistarter.org/.

SciStarter. 2023b. "SciStarter API." Accessed August 30, 2023. https://scistarter.org/API.

Scivil. 2023. "Home page." Accessed August 30, 2023. https://www.scivil.be/.

Sensor.Comunity. 2023a. "Home page." Accessed August 30, 2023. https://sensor.community/en/.

Sensor.Comunity. 2023b. "Archive." Accessed August 30, 2023. https://archive.sensor.community/.

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.

Co-funded by the European Union, Switzerland and the United Kingdom

SOSA. 2017. "Semantic Sensor Network Ontology." Accessed August 30, 2023. https://www.w3.org/TR/vocab-ssn/.

SpeciesAnalyst. 2001. "Technology Reports: The Species Analyst Project." Accessed August 30, 2023. http://xml.coverpages.org/speciesAnalyst.html.

SPOTTERON. 2023. "SPOTTERON Citizen Science home page." Accessed August 30, 2023. https://www.spotteron.net/

TagInfo. 2023. "TagInfo." Accessed August 30, 2023. https://taginfo.openstreetmap.org/.

TDWG. 2023. "Darwin Core." Accessed August 30, 2023. https://dwc.tdwg.org/.

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

University of Coïmbra. 2023. "Exploring VGI." Accessed August 30, 2023. https://vgi.uc.pt/vgi/osm/osm2lulc/.

Ústav Geoniky. 2021. "Home page." Accessed August 30, 2023. https://www.citizenscience.cz/.

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/.

VERA. 2023a. "Enabling collaborative research with and for society." Accessed August 30. 2023. https://vera.operas-eu.org/.

VERA. 2023b. "Projects." Accessed August 30, 2023. https://vera.operas-eu.org/en/search-projects.

WeObserve. 2023. "Home page". Accessed August 30, 2023. https://www.weobserve.eu.

WHO. 1980. "Glossary on air pollution. World Health Organization." Accessed August 30, 2023. https://iris.who.int/handle/10665/272866

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T. and Vieglais, D. 2012. "Darwin Core: an evolving community-developed biodiversity data standard". PloS one 7 (1): e29715.

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.

Zenodo. 2023a. "Home page." Accessed August 30, 2023. https://zenodo.org/.

Zenodo. 2023b. "Home page." Accessed August 30, 2023. https://zenodo.org/communities/geo-wiki/.

Zentrumfuercitizenscience. 2023. "Home page." Accessed August 30, 2023. https://zentrumfuercitizenscience.at/en/.

Co-funded by the European Union, Switzerland and the United Kingdom

Zooniverse. 2023. "Welcome to the Zooniverse: People-powered research." Accessed August 30, 2023. www.zooniverse.org/.