crops

Deliverable 4.2

Guidelines for citizen science data interoperability

31 December 2025

Inian Moorthy
International Institute for Applied Systems Analysis (IIASA)

Funded by
the European Union

Document details	
Project Acronym / Name	CROPS: Curating, Replicating, Orchestrating, and Propagating Citizen Science across Europe
Project URL	http://crops-cs.eu
Project Type	Coordination and Support Action (CSA)
EU Call	HORIZON-WIDERA-2023-ERA-01
Grant Agreement No.	101131696
Project Start Date	1 January 2024
Project end date	31 December 2026
Work Package	WP4 - Orchestration
Deliverable	D4.2: Guidelines for citizen science data interoperability
Due date of Deliverable	31/12/2025
Actual Submission date	30/12/2025
Lead Beneficiary for this deliverable	IIASA
Reviewed by	INOVA+
Revision	3.0
Dissemination level	Public
Number of pages	47

Document history			
Version	**Date**	**Comment**	**Modification made by**
1.0	02/12/2025	First draft	Inian Moorthy (IIASA)
2.0	05/12/2025	Minor revisions	Inês Francisco (INOVA+), Kaushiki Das (EarthWatch), Dilek Fraisl (IIASA) Inian Moorthy (IIASA)
3.0	29/12/2025	Final version	Inian Moorthy (IIASA)

To cite this document:

Moorthy, I. (2025). D4.2: Guidelines for citizen science data interoperability. *Deliverable report of project Horizon Europe CROPS (grant agreement No. 101131696)*

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Glossary and abbreviations

API	Application Programming Interface
CARE	CARE Principles for Indigenous Data Governance
CC0	Creative Commons Zero
CC BY	Creative Commons Attribution
CMEMS	Copernicus Marine Service
CRS	Coordinate Reference System
CS	Citizen Science
DCAT-AP	Data Catalogue Vocabulary – Application Profile for Europe
DOI	Digital Object Identifier
DRMKC	Disaster Risk Management Knowledge Centre
EGA	European Genome-phenome Archive
EIF	European Interoperability Framework
EMODnet	European Marine Observation and Data Network
EO	Earth Observation
EOSC	European Open Science Cloud
ERA	European Research Area
ESDAC	European Soil Data Centre
EXIF	Exchangeable Image File Format
FAIR	Findable, Accessible, Interoperable, Reusable
GBIF	Global Biodiversity Information Facility
GDC	Genomic Data Commons
GDPR	General Data Protection Regulation
GeoJSON	Geographic JavaScript Object Notation
GIS	Geographic Information System
GKH	GEO Knowledge Hub
GLEON	Global Lake Ecological Observatory Network

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

GRDC	Global Runoff Data Centre
INSPIRE	Infrastructure for Spatial Information in Europe
IoT	Internet of Things
ISO	International Organization for Standardization
ISO 19115	ISO 19115 Geographic Information Metadata
ISO 8601	ISO 8601 Date and Time Format
JSON	JavaScript Object Notation
KM4City	Knowledge Model for Smart City Ecosystem
LUCAS	Land Use/Cover Area Frame Survey
MDA	Marine Data Archive
MSFD	Marine Strategy Framework Directive
NetCDF	Network Common Data Form
OBIS	Ocean Biodiversity Information System
ODbL	Open Database License
OGC	Open Geospatial Consortium
OGC SWE	Open Geospatial Consortium Sensor Web Enablement
OSM	OpenStreetMap
PANGAEA	PANGAEA Data Publisher for Earth and Environmental Science
PRO-CTCAE	Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events
PROMIS	Patient-Reported Outcomes Measurement Information System
PPSR-Core	Public Participation in Scientific Research – Core
TCIA	The Cancer Imaging Archive
WGS84	World Geodetic System 1984

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table of Contents

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

List of Tables

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

This deliverable (D4.2) provides practical, operational guidelines for achieving interoperability and FAIR alignment in citizen science data so that it can be effectively connected, discovered and reused across European research, policy, and mission-oriented infrastructures. Building on the repository analysis conducted in D4.1 and the broader CROPS objective of supporting the upscaling of citizen science to the ERA level, this report translates conceptual interoperability requirements into concrete steps that practitioners can implement throughout the data lifecycle.

Citizen science already demonstrates strong potential to complement institutional datasets across the EU Missions; however, D4.1 highlighted persistent challenges that inhibit systematic reuse, such as inconsistent metadata, unclear licensing, heterogeneous formats, and limited pathways into mission-specific repositories. D4.2 responds directly to these gaps by providing a structured approach to designing, managing, preparing, and publishing interoperable datasets. The guidelines consolidate best practices from European and international standards including PPSR-Core, DCAT-AP, Darwin Core, INSPIRE, ISO 19115, and EOSC recommendations.

A central contribution of this deliverable is a practical, mission-agnostic interoperability checklist that helps citizen science projects make informed decisions from project planning through to publication. It emphasises four core pillars: (1) defining data and metadata requirements early, (2) maintaining methodological and structural consistency during collection, (3) documenting and validating data for reuse, and (4) selecting appropriate repositories with clear licensing and discoverability features. These principles are complemented by mission-specific data pathway analyses in Chapter 4, which illustrate how different domains intersect with European data ecosystems.

Overall, the deliverable serves as an actionable reference for citizen science practitioners, data managers and researchers. It aims to increase the visibility, quality and long-term impact of citizen science data, strengthening its contribution to the Horizon Europe Missions and to broader scientific and policy ecosystems.

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

1 Introduction

1.1 Background and purpose

The CROPS project (Curating, Replicating, Orchestrating and Propagating Citizen Science across Europe) supports the transition of citizen science (CS) from local initiatives to transnational, mission-aligned ecosystems. Within this vision, Work Package 4 (Orchestration) focuses on maximising the uptake and sustainability of citizen science activities when upscaling to the European Research Area (ERA) level.

D4.2 contributes to this goal by providing actionable guidance on how to design interoperable, FAIR-aligned and scalable dataflows that enable citizen science data to be connected, discovered and reused across European data infrastructures.

Building on the findings of D4.1 – Review of Open Data Repositories, this document translates insights into practical recommendations. D4.1 showed that Europe hosts a wide range of technically mature and FAIR-aligned repositories (i.e. Copernicus, EMODnet, GBIF, and ESDAC) but that direct integration of citizen science data remains uneven across domains. While platforms like GBIF and iNaturalist already demonstrate strong citizen participation, most EU-Mission-aligned repositories still lack mechanisms or metadata frameworks to routinely ingest and harmonise citizen-contributed datasets. Differences in metadata practices, inconsistent licensing, and limited citizen submission channels all restrict the reuse and visibility of data produced by citizens.

This deliverable also builds on earlier CROPS work on citizen science upscaling, in particular the upscaling framework and project-level assessments developed in D2.1 and D2.2, by translating identified barriers and enabling factors into practical guidance on data interoperability and reuse.

D4.2 responds to these challenges by outlining how practitioners can structure data collection, validation and publication workflows so that they align with established European and global standards. The guidelines emphasise the adoption of interoperable formats, clear metadata, open licensing and ethical data governance, ensuring that citizen science evidence can contribute meaningfully to EU Mission goals and wider policy frameworks.

1.2 The importance of interoperability for upscaling

Citizen science generates rich environmental, health and social data, yet their value is significantly reduced when held in disconnected silos. Interoperability ensures that these dispersed observations can be exchanged, interpreted and reused, transforming local contributions into evidence that supports Europe-wide analysis and decision making. Interoperable dataflows allow:

Funded by the European Union
Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

- **Integration across domains and borders.** Data from citizen science projects can feed into continental repositories supporting the five EU Missions.
- **Transparency and reproducibility.** Shared standards and metadata make methodologies and results traceable, strengthening scientific credibility.
- **Scalability.** Common architectures enable projects to evolve from local pilots to large, federated networks without losing data quality or context.
- **Sustainability.** Alignment with established infrastructures such as the European Open Science Cloud (EOSC) ensures long-term hosting, visibility and reuse.
- **Ethical and responsible practices.** The use of harmonized protocols and standards help ensure that data handling respects privacy, consent, and fair participation, supporting trustworthy, inclusive and ethical practices.

Insights from D4.1 confirm that interoperability is both a technical and organisational issue. While many repositories follow FAIR and INSPIRE guidelines, most lack mechanisms to incorporate citizen science data directly. Metadata schemas differ between missions; reuse licences are unclear; and APIs are not consistently implemented. Addressing these gaps is essential if citizen science is to contribute fully to European research, innovation and policy agendas.

1.3 Objectives and scope

This deliverable sets out guiding principles for achieving interoperability and FAIR alignment in citizen science data, supporting their integration into European research and innovation systems. It provides a strategic framework, not a technical manual, outlining how citizen science data can connect with established infrastructures and standards.

Building on insights from D4.1 – Review of Open Data Repositories, the report explains what interoperability means for citizen science, why it is essential for upscaling, and how it can enhance the visibility, credibility and reuse of citizen-contributed information.

More specifically, D4.2 aims to:

- **Clarify the key components of interoperability** in citizen science - technical, semantic, organisational and legal.
- **Summarise relevant frameworks and standards** (e.g. PPSR-Core, DCAT-AP, Darwin Core, ISO 19115) and how they apply to participatory data.
- **Highlight options for aligning citizen science data** with major European repositories and infrastructures.
- **Promote FAIR and CARE data practices** that strengthen long-term sustainability and policy relevance.

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

The deliverable serves as a reference for practitioners, data providers and policymakers, helping to ensure that citizen science contributes effectively to the European Open Science ecosystem and the Horizon Europe EU Missions.

1.4 Intended audience and use

These guidelines are relevant to a wide community involved in the collection, management and use of citizen science data across Europe. They are particularly intended for:

- **Citizen science practitioners,** who need to consider data quality, interoperability and sustainability when scaling their initiatives.
- **Data managers and repository operators**, working to align citizen-contributed datasets with existing metadata and infrastructure standards.
- **Research organisations, networks and infrastructures**, integrating participatory data into established scientific and policy frameworks.
- **Public authorities and funding bodies**, supporting the use of open and FAIR data within mission-oriented research and innovation.

For these groups, D4.2 helps bring together existing approaches and lessons learned into a coherent overview of interoperability in the citizen science context. It aims to support consistent decision-making, and a shared understanding of how citizen science data can connect more effectively with the broader European Open Science environment.

2 Understanding interoperability in citizen science

2.1 Defining interoperability

In the context of citizen science, interoperability is the ability of systems, datasets and organisations to work together, so that data collected by citizens can be understood, reused and combined with other data. Interoperability goes beyond technical compatibility; it also requires a shared understanding of what data represent, how they are described, and under what conditions they can be accessed and reused.

Citizen science in Europe is grounded in widely recognised standards of good practice, most prominently the ***ECSA Ten Principles of Citizen Science***[1] . These principles emphasise scientific value, meaningful participation, transparent data practices, open access to results where possible, ethical responsibility, and proper acknowledgement of contributors. Several principles are directly relevant to interoperability: the commitment to open sharing of data and metadata where ethically and legally appropriate (Principle 10); clear documentation of methods, processes and data creation (Principle 4); recognition and provenance of citizen contributions (Principle 7); and evaluation of data quality and project outcomes

[1] https://www.ecsa.ngo/10-principles/

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

(Principle 8). These principles shape how citizen-generated data should be curated, shared and governed. They provide the normative foundation for the technical, semantic, organisational and legal interoperability requirements adopted across Europe, ensuring that citizen science datasets are trustworthy, reusable and aligned with broader open-science expectations.

The ECSA principles set out why openness, documentation and ethical data practices are essential in citizen science. The **European Interoperability Framework (EIF)**[2] provides a *structural framework* that operationalises these expectations across European data systems. Together, they clarify both the motivation for interoperability and the mechanisms through which it is achieved.

According to the European Interoperability Framework (EIF), four complementary dimensions are typically recognised:

Technical interoperability – the use of compatible data formats, open APIs, identifiers and exchange protocols that enable data to move reliably between systems.

Semantic interoperability – ensuring that the meaning of data is preserved through shared vocabularies, ontologies and metadata standards.

Organisational interoperability – coordination of processes, roles and workflows between projects, repositories and infrastructures so that data exchange is consistent and sustainable.

Legal interoperability – clarity around data ownership, licensing, consent and privacy, ensuring that information can be shared and reused lawfully and ethically across jurisdictions.

These dimensions provide a structured way to understand and address the multiple layers that influence whether citizen science data can effectively interact with institutional and policy data systems. In practical terms, interoperability creates the bridge between local observations made by citizen scientists and the structured data ecosystems that underpin European research and innovation. It allows citizen science outputs to become part of a wider body of open, verifiable evidence rather than stand-alone community records.

2.2 Interoperability and the FAIR principles

The **FAIR principles (Findable, Accessible, Interoperable and Reusable)** provide a common framework for open data management across Europe. They underpin the European Open Science Cloud (EOSC) and guide how citizen science data can be made visible and reusable beyond individual projects.

[2] European Commission: Directorate-General for Digital Services, *New European interoperability framework – Promoting seamless services and data flows for European public administrations*, Publications Office, 2017, https://data.europa.eu/doi/10.2799/78681

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

For citizen science, FAIR means creating data practices that allow community observations to connect with research and policy systems:

Findable: Data are described with clear metadata, persistent identifiers and searchable catalogue entries.

Accessible: Data and metadata can be retrieved through standard, open protocols with transparent access conditions.

Interoperable: Information follows recognised vocabularies and models such as PPSR-Core, DCAT-AP or Darwin Core so it can be combined with other datasets.

Reusable: Data include context on methods, quality and licensing to support future use.

Applying FAIR remains uneven across citizen science. D4.1 found that while most major European repositories comply with FAIR, citizen science datasets often lack consistent metadata, machine-readable formats or clear reuse licences.

FAIR provides both a goal and a practical guide. Incremental improvements, using open formats, adopting shared metadata fields and clarifying licences, can greatly increase the discoverability and long-term value of citizen science data.

While FAIR focuses on the technical and semantic qualities that make data reusable, some citizen science contexts also require attention to equity, ethics and community rights. The **CARE Principles for Indigenous Data Governance**[3] (*Collective Benefit*, *Authority to Control*, *Responsibility* and *Ethics*) complement FAIR by highlighting the social dimensions of responsible data stewardship. CARE is not universally applicable across all citizen science domains, but it is relevant where data relate to communities, traditional knowledge or sensitive contexts, helping ensure that interoperability supports both technical reuse and ethical governance.

2.3 The European interoperability landscape for citizen science data

Citizen science data operate within a wider European ecosystem shaped by policy, technical standards, and shared data infrastructures. For many projects, this landscape can seem complex. To support clarity and practical decision making, we present it through three consolidated layers:

1. **Policy and governance;**
2. **Standards and technical requirements;**
3. **European data infrastructures.**

[3] https://www.gida-global.org/care

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

This structure helps practitioners understand why interoperability matters, how to implement it, and where their data can ultimately be shared or reused.

2.3.1 Policy and Governance Frameworks

These frameworks define the **principles and strategic direction** for how data should be shared and managed in Europe. They do not prescribe specific formats, but they establish the expectations for openness, FAIRness, and cross-sector collaboration.

European Interoperability Framework (EIF)[4]: The EIF provides a common reference for interoperability across the European Union. It defines four complementary dimensions (technical, semantic, organisational, and legal) that support data exchange and cooperation between systems and institutions. For citizen science, these dimensions help projects structure their data and governance practices in ways that make collaboration and reuse possible across national and sectoral boundaries.

European Open Science Cloud (EOSC)[5]: EOSC is a federated environment that provides access to FAIR and open data, computing resources, and related services for research and innovation. It serves as Europe's main infrastructure for sharing and accessing scientific data across disciplines. Aligning citizen science repositories with EOSC principles improves visibility, ensures long-term access, and connects citizen science data with formal research outputs.

COST Action CA15212[6]: A European network that has advanced standardisation and interoperability practices in citizen science. It provides a collaborative framework for improving metadata standards, data management, and ethical practices. Its outputs inform how CROPS and similar initiatives approach data integration and community engagement.

GDPR[7]: For datasets that include personal or potentially identifiable information, the General Data Protection Regulation (GDPR) also shapes how FAIR and interoperable practices can be applied. GDPR does not prevent interoperability, but it requires that access conditions, consent, licensing and data documentation are managed in ways that protect individuals while still enabling responsible reuse.

Together, these frameworks explain why interoperability is needed and establish a shared vision for consistent, FAIR data across Europe. In addition to these policy and governance frameworks, several EU-funded projects have explored practical approaches to improving interoperability in citizen science. Initiatives such as COS4CLOUD have examined how shared services, metadata alignment, and

[4] European Commission: Directorate-General for Digital Services, *New European interoperability framework – Promoting seamless services and data flows for European public administrations*, Publications Office, 2017, https://data.europa.eu/doi/10.2799/78681
[5] European Commission: Directorate-General for Research and Innovation, *Realising the European open science cloud – First report and recommendations of the Commission high level expert group on the European open science cloud*, Publications Office, 2016, https://data.europa.eu/doi/10.2777/940154
[6] https://cs-eu.net/
[7] https://gdpr-info.eu/

Funded by the European Union
Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

integration with European research infrastructures can support the reuse of citizen-generated data. The lessons from these projects highlight both the potential and the challenges of harmonising citizen science data across domains and provide useful contextual insight for the guidance developed in this deliverable.

2.3.2 Standards and Technical Requirements

This layer contains the practical specifications and regulatory frameworks that determine how data should be structured, described, and exchanged. These are the tools citizen science projects use to make their data usable beyond their immediate context.

Regulatory and harmonisation frameworks

INSPIRE Directive[8]: Defines harmonised structures for geospatial and environmental data. Many datasets relevant to the EU Missions (e.g., soil, water, biodiversity, climate impacts) fall within INSPIRE themes. Alignment with INSPIRE concepts improves discoverability and integration with public authorities.

Open Data Directive[9]: Sets rules for re-use of public-sector information, emphasising openness, machine-readability, and use of standard APIs.

General metadata standards

DCAT-AP[10]**:** The Data Catalogue Application Profile for Europe (DCAT-AP) defines how datasets and catalogues are described on open data portals. It supports metadata exchange between national and European catalogues, allowing citizen science datasets published through local or institutional portals to be discovered through larger EU data infrastructures.

Dublin Core[11]**:** A flexible, multi-domain vocabulary widely used by many repositories.

Citizen science–specific standards

PPSR-Core[12]**:** PPSR-Core is an international metadata standard developed specifically for citizen science. It provides structured fields to describe projects, datasets, and individual observations in a machine-readable way. Using PPSR-Core helps citizen science initiatives make their data compatible with repositories such as GBIF and the European Citizen Science Platform, supporting cross-project discovery and reuse.

[8] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) http://data.europa.eu/eli/dir/2007/2/oj
[9] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast) http://data.europa.eu/eli/dir/2019/1024/oj
[10] https://op.europa.eu/de/web/eu-vocabularies/dcat-ap
[11] https://www.dublincore.org/resources/userguide/
[12] https://core.citizenscience.org/

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Domain standards

Darwin Core[13]: Darwin Core is a global standard for sharing biodiversity data. It specifies a simple vocabulary for describing species occurrences and related observations. The standard underpins platforms such as GBIF and iNaturalist, where citizen science data are routinely integrated with professional biodiversity monitoring efforts.

ISO 19115[14]: International standard for defining the metadata schema required to describe geographic information and associated services. Its primary purpose is to enable users to effectively discover, assess the suitability of, and properly utilize digital geographic datasets by providing comprehensive information about their identification, quality, extent, and distribution.

These standards define how data should be described and formatted so that it can be integrated across repositories, models, and mission data ecosystems. While these frameworks and standards provide the foundation for interoperability, their practical application can vary across citizen science domains. To support consistent implementation, Table 1 summarises how the most relevant standards map onto typical citizen science data types and use cases. This provides a concise reference for practitioners when selecting appropriate standards for designing, documenting and publishing interoperable datasets.

Table 1: Mapping of key interoperability standards to typical citizen science use cases.

Standard / Framework	What It Provides	How Citizen Science Projects Use It (Typical Applications)	Example CS Domains
PPSR-Core	Structured project- and dataset-level metadata for participatory science	· Describe project purpose, participant roles, protocols, and dataset lineage· Provide consistent metadata for discovery across platforms· Harmonise project descriptions for portals	Cross-cutting; general CS platforms; social science; health behaviour
DCAT-AP	Harmonised dataset metadata for European data catalogues	· Publish CS datasets on national/EU open data portals· Ensure datasets appear in federation mechanisms· Improve discoverability outside domain-specific repositories	Climate, cities, soil, water, environmental observations
Dublin Core	Flexible general-purpose metadata vocabulary	· Provide minimal descriptive metadata for simple CS datasets· Use for portals and repositories requiring lightweight metadata·	Social science CS projects; community reporting

[13] https://dwc.tdwg.org/
[14] https://www.iso.org/obp/ui/es/#iso:std:iso:19115:-1:ed-1:v1:en

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Standard / Framework	What It Provides	How Citizen Science Projects Use It (Typical Applications)	Example CS Domains
		Apply for qualitative or mixed-method datasets	
Darwin Core	Terms for species occurrence and biodiversity data	· Structure species observations· Ensure compatibility with GBIF and biodiversity platforms· Provide taxonomic, temporal and spatial fields	Biodiversity, phenology, coastal and marine species monitoring
INSPIRE (themes + vocabularies)	EU-standard geospatial metadata, environmental parameters, CRS definitions	· Structure geospatial attributes (CRS, geolocation, accuracy)· Align environmental variables (soil, water, air)· Ensure compatibility with environmental agencies	Soil health, water quality, climate adaptation, land cover, erosion
ISO 19115	International geospatial metadata standard	· Provide detailed geospatial metadata for datasets intended for modelling, GIS or long-term archiving	Hydrology, cryosphere, coastal mapping, land-use studies
OGC SensorThings API[15]	Standard for real-time and IoT sensor data	· Publish sensor data streams from low-cost devices· Encode time series with sensor ID, calibration and frequency· Enable integration with municipal dashboards and modelling tools	Air quality, noise, temperature, soil moisture, water levels
OGC SWE (Sensor Web Enablement)[16]	Framework for interoperability of sensors	· Document metadata for environmental sensors· Standardise sensor observations for hydrology or marine monitoring	Hydrology, marine sensors, lake observatories
EXIF (image metadata)	Standard metadata properties in photos	· Provide timestamp, location, orientation for CS images· Support automated extraction for mapping and CV models	Urban imagery, coastal erosion, pollution reporting
OSM Tagging Schema	Controlled vocabulary for geographic features	· Standardise tagging of urban features (benches, crossings, trees)· Ensure compatibility with OSM and municipal datasets	Cities, mobility, accessibility audits, public-space mapping

[15] https://www.ogc.org/standards/sensorthings/
[16] https://www.ogc.org/standards/swes/

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Standard / Framework	What It Provides	How Citizen Science Projects Use It (Typical Applications)	Example CS Domains
PRO-CTCAE / PROMIS[17]	Patient-reported outcome vocabularies	· Standardise symptom and outcome reporting· Enable comparability of cancer-related CS datasets	Health & cancer citizen science (non-clinical)

3 Designing interoperable and FAIR citizen science data flows

Citizen science data become interoperable and mission-ready when projects plan for reuse from the start and follow a clear, structured data lifecycle. This chapter presents four practical steps that any project can adopt: **Plan**, **Manage**, **Prepare**, and **Publish**.

- **Plan**: Define data formats, metadata fields, identifiers, and vocabularies before collection begins, drawing on recognised standards such as PPSR-Core, DCAT-AP, INSPIRE, Darwin Core, or domain-specific schemas where relevant.
- **Manage**: Coordinate contributors and maintain data quality during collection by ensuring consistent units, sensor calibration notes, timestamps (ISO 8601), and geolocation (WGS84), while updating metadata continuously.
- **Prepare**: Consolidate, clean, and document datasets for reuse by adding a data dictionary, describing processing steps, and aligning variables with recognised standards or controlled vocabularies.
- **Publish**: Select an appropriate repository, provide clear metadata, apply an open license, and register a persistent identifier so datasets can be discovered, harvested, and reused by platforms identified in D4.1.

These four steps provide a simple, repeatable structure for producing FAIR-aligned and interoperable citizen science datasets that can connect to European infrastructures and Mission data ecosystems, while reflecting the core ECSA principles.

3.1 Plan: Define what you will collect and how

Interoperability depends on decisions made at the earliest stages of a project. When citizen science initiatives begin collecting data before defining structure, formats, or documentation standards, inconsistencies quickly accumulate and become difficult to correct. Establishing a clear data design from the outset

[17] https://healthcaredelivery.cancer.gov/pro-ctcae/overview.html

Funded by the European Union
Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

ensures that citizen science datasets can be integrated, validated, and reused within wider scientific, policy, and repository ecosystems.

This section outlines the essential elements that practitioners should define before data collection begins.

3.1.1 Data collection

Citizen science data should be collected in open, machine-readable formats such as CSV, JSON, GeoJSON, or for gridded data NetCDF. These formats align with European repository expectations and can be processed by standard analytical tools. Avoid proprietary spreadsheets with formulas or embedded objects, which limit reuse.

To ensure consistency across contributors:

- Use controlled vocabularies or predefined options for core variables; minimise free text.
- Choose domain-aligned schemas where relevant, such as Darwin Core for biodiversity, INSPIRE vocabularies for environmental parameters, or the OGC SensorThings model for sensor streams.

Define persistent identifiers for sampling sites, sensors, features, or species. IDs do not need to be complex but must remain stable across datasets, enabling temporal linkage and integration with institutional records.

Each observation should always include:

- Timestamp (ISO 8601)
- Location (WGS84 coordinates or administrative unit)
- Method or protocol used
- Device, sensor, or app version

These elements form the minimum contextual information required by most repositories.

3.1.2 Metadata essentials

Metadata describe what a dataset contains, how it was produced, and how it should be interpreted. In citizen science, where observations may be contributed by volunteers with varying levels of expertise and across different locations and time periods, metadata are crucial for ensuring that datasets remain understandable and usable. Without sufficient metadata, datasets often lose scientific value, become difficult to integrate with other sources, or cannot be published in established repositories.

Creating high-quality metadata is not only a documentation exercise. It is an essential part of designing a data flow that supports transparency, reproducibility,

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

and interoperability. Practitioners should treat metadata as a permanent companion to the dataset, developed at project outset, updated during data collection, and finalised at publication.

Metadata for citizen science projects can be organised into four core categories: descriptive, methodological, spatial and temporal, and structural metadata. Each category plays a specific role in making data discoverable, comparable, and reusable.

Descriptive metadata

Descriptive metadata provide the high-level context that allows users to understand what the dataset is about. They should enable someone who has never participated in the project to grasp the purpose of the data collection within minutes. Typical descriptive metadata include:

- the title of the dataset
- a concise overview of the topic, scope, definitions, concepts and phenomenon being observed
- the scientific or monitoring objective
- the responsible organisation or project leader
- links to project documentation, websites, protocols, or dashboards

This information supports discoverability in catalogues and helps repositories classify the dataset correctly.

Methodological metadata

Methodological metadata describe how the data were created. In citizen science, where methods may be simple or distributed across many contributors, these details are essential for assessing data quality. They should include:

- a clear description of the sampling or reporting protocol
- instructions provided to volunteers
- the instruments, sensors, or software used, including versions
- calibration routines or performance checks
- how the data are collected
- criteria for validating or rejecting observations
- quality assurance and control mechanisms and processes
- any known methodological limitations

If the protocol changed during the project, the metadata should record when and why. Even minor adjustments, such as switching to a new app version or updating a sensor calibration method, can affect data comparability.

Spatial and temporal metadata

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Spatial and temporal information is often the most important dimension for reuse. Missing or inconsistent location or time information is one of the most common barriers to dataset integration. Spatial and temporal metadata should specify:

- coordinates or location descriptors
- the coordinate reference system (for example WGS84)
- spatial precision (for example exact GPS point, 100 m grid cell, or municipality)
- methods for capturing location (GPS, manual input, fixed sensor)
- observation timestamps in a consistent format (for example ISO 8601)
- the overall temporal coverage of the dataset
- observation frequency or sampling interval

Clear spatial and temporal metadata allow datasets to be used by modelling tools, linked with remote sensing products, or aligned with institutional datasets.

Structural metadata

Structural metadata describe how the dataset is organised internally. They explain what each column or field represents and how the dataset should be interpreted. Structural metadata include:

- variable names and detailed definitions
- units of measurement
- valid ranges or permissible values (for example species codes, pollutant classes, weather types)
- data types (for example numeric, categorical, boolean, text)
- relationships between tables if the dataset includes multiple files
- conventions for missing or null values
- the file structure and naming conventions

A clear data dictionary that lists each variable and its meaning is one of the most powerful enablers of interoperability. It prevents misinterpretation and makes it easier for others to map the dataset to standards such as Darwin Core, ISO 19115, or DCAT-AP.

Metadata as a living component of the project

Metadata should not be created only at the end of a project. They work best when maintained throughout the data lifecycle. Practitioners should update metadata whenever protocols evolve, new instruments are introduced, or validation processes change. Keeping metadata current avoids inconsistencies and ensures that the final dataset reflects what actually happened during data collection.

Standardised metadata templates

Using a standardised metadata template is one of the simplest ways to ensure completeness and consistency. A template acts as a checklist that prevents critical

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

information from being overlooked. It also supports training, reduces the burden on contributors, and allows technical partners to process the data more efficiently. projects may incorporate elements from:

- PPSR-Core (project metadata)
- DCAT-AP (dataset-level discovery metadata)
- Darwin Core (biodiversity observations)
- INSPIRE/ISO 19115 (spatial and environmental data)

Templates ensure harmonisation and reduce the burden of preparing datasets for publication. Projects may also incorporate elements from recognised standards depending on their domain and repository requirements.

3.2 Manage: Maintain consistency and data quality during collection

Once data collection begins, the focus shifts from planning to the ongoing task of maintaining consistency and quality. Citizen science projects may gather data over long periods and across many contributors, tools, and locations. Without active coordination, small inconsistencies can accumulate and undermine the coherence of the dataset. Managing data during collection ensures that contributions remain aligned with the original design, that issues are identified early, and that the final dataset can be prepared for reuse without major repairs.

This section provides practical guidance on how to coordinate contributors, monitor data quality, and keep both data and metadata consistent as the project evolves.

3.2.1 Maintain consistency in how data are collected

After data collection begins, practitioners should periodically confirm that incoming records still follow the structure and conventions defined in **Section 3.1**. In citizen science, even small deviations can multiply across many volunteers.

Regular checks should verify that:

- required fields are being completed
- units and classifications remain consistent
- identifiers (e.g., site IDs, sensor IDs) are used reliably
- timestamps use **ISO 8601** and locations follow **WGS84** or the agreed reference system
- tools or apps have not introduced unexpected changes (e.g., updates that alter field names or values)

Short, periodic reviews are often sufficient to detect misunderstandings, unclear instructions, or interface issues. Clear communication with volunteers such as reminders, small corrections, and example submissions helps maintain alignment.

Opportunities for contributors to ask questions or receive feedback also strengthen accuracy and confidence.

3.2.2 Monitor data quality and address issues as they arise

Throughout the data collection phase, practitioners should monitor the quality of incoming data and record any issues that may affect interpretation. Early detection allows problems to be addressed while they are still manageable.

Useful practices include:

- running basic automated checks to flag missing fields or unexpected values
- comparing new records with earlier data to identify shifts or anomalies
- checking distributions of key variables for sudden changes
- reviewing observations from new contributors or new tools
- tracking data volumes to identify gaps in time or space

When issues appear, practitioners should document the cause and the steps taken in response. For example, if a sensor calibration drift is discovered, the affected period should be noted in the metadata. If volunteers misinterpret a category, instructions can be adjusted, and the point of change should be recorded.

Citizen science projects often evolve over time. Applications may update, new volunteers may join, or additional organisations may contribute data. Each of these changes may affect consistency. It is important to document when and why these changes took place and to communicate them clearly to contributors. This helps ensure that subsequent users of the dataset can understand variations that appear within the data.

3.2.3 Organise data and metadata throughout the collection process

Good data management during collection reduces complexity later. Practitioners should maintain clear and secure storage systems that separate raw data from working or cleaned files. Raw data should never be overwritten, and all changes should be logged. Even simple version control, such as maintaining dated folders or change logs, helps maintain transparency and reproducibility.

In parallel, metadata should be updated as the project progresses. This may include changes to field definitions, updates to equipment, refinements to the protocol, or clarifications based on volunteer feedback. Keeping metadata current ensures that the final dataset accurately reflects how data were produced and reduces the likelihood of missing or incorrect documentation at publication time.

Maintaining both data and metadata in an organised manner supports the transition to the next stage, where datasets are prepared for publication and reuse.

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

It also reduces the risk of data loss, confusion, or inconsistencies that could affect long-term value.

3.3 Prepare: Clean, Document and Validate data for reuse

Citizen science generates valuable environmental and societal data, but these datasets only achieve their full potential when they can be understood and reused beyond the original project. As noted in Chapters 1 and 2, interoperability and FAIR alignment are key to enabling citizen science contributions to support research, policy, and EU Mission objectives. Preparing datasets for publication is therefore not simply an administrative step. It is the moment where the results of community engagement are transformed into a resource that can be integrated into larger scientific and policy infrastructures such as EOSC, GBIF, EMODnet, or data.europa.eu.

This section explains how practitioners can prepare a coherent and complete dataset package that reflects the realities of citizen science data collection and meets the expectations of external users and repositories.

3.3.1 Consolidate and review the dataset

Citizen science projects often collect data through a combination of mobile apps, online forms, sensors, social media integrations, partner organisations, and in some cases, paper-based records. This diversity is a strength, but it also means that data arrive in different formats and levels of completeness.

Before publication, practitioners should bring all contributions together and confirm that they form a coherent dataset. This includes checking that:

- records collected across different tools follow the expected structure
- required fields defined during project design (Section 3.1) are present
- observations include the contextual details needed for interpretation (e.g., ISO 8601 timestamps, WGS84 locations)
- sensor or app-based data align with the metadata created during collection (Section 3.2)

If datasets are incomplete or inconsistently structured at this stage, many repositories will be unable to accept or reuse them. Consolidation therefore provides a final opportunity to ensure that the dataset is complete, well-documented, and aligned with the project's intended design.

3.3.2 Describe how the data were processed

Citizen science datasets often undergo several rounds of cleaning and transformation. These may include correcting species names, validating sensor readings, adjusting timestamps, or reconciling duplicate entries from multiple volunteers at the same location. As highlighted in Chapter 2, transparency in processing is central to FAIR principles and is essential for downstream reuse.

A short narrative describing these steps helps external users assess reliability and understand limitations. This narrative should explain:

- how data collected by volunteers were validated
- how sensor or model outputs were checked or calibrated
- how obvious errors or inconsistencies were corrected
- whether expert review or automated validation was used
- which parts of the dataset may require caution or additional filtering

This information ensures that users can judge whether the dataset is suitable for their needs, whether for research, education, or integration into broader monitoring systems.

3.3.3 Prepare clear descriptions of variables

Fields that are obvious to project participants may be unclear to external users. A variable description table (data dictionary) ensures that each field is understandable and correctly interpreted. A clear table that describes each variable helps bridge this gap. This table should explain:

- how variables relate to the underlying phenomenon being monitored
- how volunteers or sensors generated these observations
- any classifications, codes, or controlled vocabularies used (e.g., Darwin Core terms, INSPIRE categories, OSM tags)
- how values should be interpreted in relation to project objectives

This practice directly supports semantic interoperability as described in Chapter 2.1, ensuring that data from different communities can be combined and understood consistently.

3.3.4 Assemble the dataset package

European repositories and research infrastructures expect datasets to be self-contained. For citizen science, this means packaging the data with the supporting material that explains how observations were created, curated, and interpreted.

A complete package typically includes:

- the cleaned dataset ready for reuse
- a link or reference to the raw data (where ethically and legally possible)
- metadata created at the start of the project and updated throughout collection
- the variable description table
- a short project summary describing purpose, methods, and participants
- a processing overview documenting major cleaning and transformation steps

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

This package ensures that the dataset can stand alone and is ready for repositories that require independent, well-documented data submissions, such as Zenodo, GBIF, EMODnet, national open data portals, or other infrastructures used by Mission actors.

3.3.5 Carry out final checks before publication

Citizen science datasets are often intended for broad reuse, including scientific research, policy development, ecosystem modelling, and public engagement. Before publication, a final review helps ensure that the dataset meets the practical requirements of these audiences.

Practitioners should confirm that:

- fields match their descriptions and metadata
- formats are consistent across the entire dataset
- timestamps and spatial information can be interpreted without ambiguity
- no unintended corrections or formatting issues remain from spreadsheets or export processes
- the dataset is internally coherent given the project's scope and objectives

Completing these checks helps ensure that citizen science data can be ingested into the European data ecosystem and support the evidence base for environmental monitoring and EU Missions, as discussed in Chapters 1 and 2.

3.4 Publish: Share data for integration and long-term reuse

Publishing a dataset is the moment when citizen science contributions enter the wider knowledge ecosystem. As outlined in Chapters 1 and 2, citizen science data can support research, policy, and European Missions when it is made visible, accessible, and linked to established infrastructures. Publication is therefore not only the final stage of a project's data workflow, but also the point where community efforts can have lasting impact.

To publish effectively, practitioners need to select a suitable repository, meet its expectations, and ensure that the dataset is represented clearly and accurately so others can use it without additional guidance.

3.4.1 Choose an appropriate repository and meet its requirements

Citizen science projects produce varied types of data, and the choice of repository (Table 2) depends on the kind of observations collected and the communities that might use them. General-purpose repositories such as Zenodo or Figshare are useful for projects with mixed data types or formats. They provide persistent identifiers and long-term access, and their flexibility suits many citizen science contexts.

For projects aligned with established scientific domains, domain-specific repositories offer stronger integration pathways. Biodiversity observations fit naturally within GBIF, where large volumes of citizen science data are already aggregated. Water and marine datasets may align with EMODnet or SeaDataNet, while soil-related observations can support ESDAC resources. Datasets intended for broad public access or public administration can be catalogued through data.europa.eu.

Each repository comes with specific expectations. Some require particular metadata structures, such as Darwin Core for species records or INSPIRE elements for spatial data. Others expect information about the provenance of observations, the roles of volunteers, or the methods used to validate contributions. Reviewing and meeting these requirements helps ensure that the dataset is accepted, indexed correctly, and integrated with related data.

Table 2: Overview of recommended repositories (from D4.1) for citizen science data across the five EU Missions.

Category/ Mission	Repositories
General	• Dryad • Zenodo • Figshare • EOCS EU Node
Adaptation to climate change	• Copernicus Climate Data Store (CDS) • European Climate Data Explorer • DRMKC Risk Data Hub • Global Biodiversity Information Facility (GBIF) • World Meteorological Organization (WMO) Data Platform
Cancer	• The Cancer Imaging Archive • PatientsLikeMe • Genomic Data Commons (GDC) • European Genome-phenome Archive (EGA)
Climate-neutral and smart cities	• Smart Cities Marketplace • Smart City Platform • Zenodo (Urban Projects) • OpenStreetMap (OSM) • KM4City Ecosystem • Panoramax
A soil deal for Europe	• European Soil Data Centre (ESDAC) • ISRIC-World Soil Information • LUCAS Soil • OpenLandMap

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Restore our oceans and waters	• EMODnet, Marine Data Archive (MDA) • SeaDataNet • iNaturalist (Marine) • Copernicus Marine Service (CMEMS) • Ocean Biodiversity Information System (OBIS) • PANGAEA • FreshWater Watch • GLEON • Global Runoff Data Centre (GRDC) • HydroShare • AquaSTAT

3.4.2 Create a clear and discoverable dataset record

Once a repository has been selected, the dataset needs to be described in a way that supports discovery and reuse. A clear, concise record helps users understand what the dataset contains, why it was collected, and how it can be interpreted. This is especially important for citizen science projects, where methods and participation models may differ from conventional scientific studies.

A strong dataset record should explain the project's purpose, the type of observations collected, the geographic or thematic scope, and the key elements of the data production process. Repositories make use of keywords, geographic descriptors, and domain-specific terms to aid discovery, so using familiar and accurate terminology improves visibility.

Most repositories assign persistent identifiers such as DOIs. These identifiers make the dataset citable and stable, and they allow it to be linked to other project materials such as training resources, dashboards, or publications. Connecting these resources strengthens transparency and shows how citizen science contributions were produced and used.

3.4.3 Support long-term access and dataset evolution

Citizen science datasets often continue to grow after their initial publication. Volunteers may remain active, new sites may be added, or additional measurements may be collected. Practitioners should consider whether the published dataset will remain static or whether updated versions will be released over time.

If updates are expected:

- maintain versioning, clearly identifying the changes made
- update metadata and variable descriptions accordingly
- create concise release notes to help users assess whether the new version meets their needs

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

- preserve access to earlier versions or raw data for transparency and reproducibility

These practices ensure that datasets remain trustworthy and usable long after their initial publication and support the sustainability principles outlined in Chapter 1.

4 Citizen Science Data Interoperability and Repository Options for the EU Missions

4.1 Why Mission alignment matters for CS data interoperability

The five EU Missions depend on high-quality data that can be compared, aggregated, and interpreted across countries, disciplines, and monitoring frameworks. They bring together public authorities, research institutions, private sector actors, and local communities to address complex challenges such as climate resilience, water restoration, sustainable soil management, urban transformation, and cancer prevention. Within this landscape, citizen science has emerged as a valuable contributor, offering observations and perspectives that institutional monitoring systems alone cannot provide.

Citizen science data are often collected at local scale, in high temporal frequency, and in places or contexts where traditional monitoring is limited. Volunteers report climate impacts as they unfold, observe local water quality issues, measure air pollution at the street level, participate in soil monitoring campaigns, and document environmental factors linked to health. These contributions enrich European datasets by adding granularity, extending coverage, and providing community-driven insights that support more inclusive decision making.

However, the usefulness of these contributions depends on their interoperability. As outlined in Chapter 3, citizen science projects must structure their data, metadata, and documentation in ways that allow other actors to interpret and reuse them. Without clear spatial references, transparent methods, and traceable provenance, data remain local observations rather than mission-relevant evidence.

The findings from D4.1 show clearly that each Mission has a distinct data ecosystem and repository landscape. Some domains, such as marine monitoring or biodiversity, have well-established infrastructures with recognised standards. Others, such as urban data or soil health, rely more on open-data catalogues or national-level systems. Integration readiness varies significantly, and direct ingestion of citizen science datasets is rare in most domains. Instead, repositories use structured metadata, standard vocabularies, and catalogue records to identify and link open datasets published elsewhere.

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

This is why interoperability is central. When citizen science datasets follow the practices described in Chapter 3 (i.e. consistent formats, clear metadata, method transparency, spatial precision, and well-documented validation) they become discoverable, reusable, and comparable within Mission workflows, even if they are not ingested directly.

Mission alignment therefore matters in two ways:

It helps practitioners understand where their data can have the most impact. Different Missions value different types of observations, and knowing the data landscape allows projects to publish strategically.

It clarifies what interoperability looks like in each domain. Climate adaptation depends on precise spatial metadata, marine monitoring relies on INSPIRE alignment, urban missions expect DCAT-AP, soil missions require detailed sampling context, and cancer-related initiatives emphasise ethical and legal clarity.

Chapter 4 provides a Mission-by-Mission overview of these data ecosystems and explains how interoperable citizen science data connect to them. The aim is not to prescribe technical compliance, but to help practitioners position their data within the European landscape in ways that maximise visibility, usability, and long-term value. In this chapter, repositories are understood broadly to include not only domain-specific platforms that accept direct data submissions but also federated catalogues and discovery services that surface interoperable datasets published elsewhere. Such platforms play an important role in enabling cross-domain visibility and reuse of citizen science data within European and international data ecosystems, even where direct ingestion into mission-specific repositories is not possible.

4.2 Mission: Adaptation to Climate Change

The Adaptation to Climate Change Mission supports European regions and communities in understanding climate risks, planning adaptation pathways, and testing solutions. Citizen science can fill important spatial and temporal gaps in climate-related information. Local biodiversity observations, snow and rainfall measurements, extreme weather reporting, land cover changes, and community-based risk assessments all complement institutional datasets. These contributions are valuable when structured in ways that enable integration into the Mission's wider data ecosystem.

From D4.1, the key adaptation-relevant repositories are:

- Copernicus Climate Data Store (CDS)
- European Climate Data Explorer

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

- DRMKC Risk Data Hub
- Global Biodiversity Information Facility (GBIF)
- World Meteorological Organization (WMO) Data Platform

Among these, only GBIF accepts direct citizen science submissions. All others rely on institutional or national pipelines. This means citizen science datasets must be published openly, documented clearly, and structured for reuse so that researchers, public authorities, and intermediaries can integrate them with climate indicators and hazard layers.

4.2.1 Mission-specific data and interoperability requirements

To ensure that citizen science observations can be combined with climate indicators, hazard layers, hydrological models, and ecological datasets, projects supporting the Adaptation Mission should meet the following cross-cutting technical requirements:

- **WGS84 geolocation** with INSPIRE-compliant spatial metadata (CRS, accuracy, location method).
- **ISO 8601 timestamps** with time zone indication and clear sampling intervals.
- **Standard environmental units**, such as mm (precipitation), °C (temperature), % (soil moisture), and cm (snow depth).
- **Relevant vocabularies**, including DRMKC hazard categories, INSPIRE environmental themes, and **Darwin Core** for biodiversity and phenology.
- **Sensor metadata** (device model, calibration method, firmware version, frequency, stable sensor ID) for weather, hydrology, or environmental sensors.
- **Machine-readable formats** (CSV, JSON, GeoJSON; **NetCDF** for gridded or model-ready outputs).
- **Method and lineage documentation**, including protocols, validation rules, and known limitations.
- **Open licences** (CC BY or CC0) enabling reuse by climate services and public authorities.

Table 3 summarises the technical requirements for common adaptation-relevant data types and how they are used by adaptation actors.

Table 3: Key citizen science data types and interoperability requirements for the Adaptation to Climate Change Mission

Type of Adaptation-Relevant CS Data	Essential Metadata & Attributes	Relevant Standards / Conventions	How Adaptation Actors Use It
Local climate impacts (heat stress, drought, soil moisture)	Coordinates (WGS84), time (ISO 8601), variable definition,	INSPIRE environmental themes; ISO 19115	Complements climate indicators

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Type of Adaptation-Relevant CS Data	Essential Metadata & Attributes	Relevant Standards / Conventions	How Adaptation Actors Use It
	units, sampling interval		and vulnerability assessments
Extreme weather observations (rainfall, hail, wind, flash floods)	Geolocation, time, event type, severity, optional photos, uncertainty flag	DRMKC hazard categories; EXIF	Supports hazard mapping and disaster-loss analysis
Hydrological data (water levels, overflow, drainage blockage)	Coordinates, timestamp, measurement method, units, sensor ID, frequency	ISO 19115; OGC SensorThings	Supports flood-risk assessments and drainage modelling
Cryosphere observations (snow depth, freeze–thaw events)	Coordinates, timestamp, method, units, slope/aspect	INSPIRE elevation/land cover	Supports melt modelling and mountain adaptation
Biodiversity & phenology (species presence, phenology stages)	Species name, event date, location, observation method	Darwin Core terms	Provides ecological indicators of climate change
Environmental quality indicators (temperature, soil moisture sensors)	Sensor model, calibration notes, frequency, units, stable ID	OGC SensorThings; INSPIRE parameters	Supports monitoring of heat stress, drought, and local climate impacts
Community vulnerability indicators (heat discomfort, water access issues)	Location (coarse if sensitive), timestamp, anonymous ID, vulnerability type	ISO 37120 categories (where relevant)	Supports contextual elements of local adaptation planning

4.2.2 Repository options for Citizen Science Data Relevant to the Adaptation Mission

Citizen science data relevant to the Adaptation to Climate Change Mission are generally not ingested directly into core European climate repositories, but can nonetheless contribute to adaptation planning and analysis when published openly and documented using recognised standards. The repository options outlined below (Table 4) describe the main pathways through which interoperable citizen science datasets can be discovered, reused, and combined with institutional climate data infrastructures identified in D4.1, including federated discovery platforms like the GEO Knowledge Hub (GKH).

Funded by the European Union
Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table 4: Repository options for citizen science datasets supporting the Adaptation to Climate Change Mission

Type of Adaptation-Relevant CS Data	Preferred Publication Route	Alternative Route	Which Adaptation Repositories Can Reuse / Compare It (Indirectly)
Local climate impacts (heat stress, drought stress, soil moisture)	National environmental or climate portals	Zenodo or GKH	CDS, Climate Explorer
Extreme weather observations (rainfall, hail, wind, flash floods)	National hazard / meteorological portals	Zenodo or GKH	CDS, WMO
Hydrological data (water levels, overflow, drainage blockage)	National hydrology or environmental portals	Zenodo or GKH	DRMKC, CDS (contextual)
Cryosphere observations (snow depth, freeze–thaw events)	National climate / cryosphere portals	Zenodo or GKH	Climate Explorer, CDS
Biodiversity & phenology (species presence, phenology stages)	GBIF (Darwin Core)	-	GBIF, Climate Explorer
Environmental quality indicators (temperature, soil moisture sensors)	National environmental portals	Zenodo or GKH	CDS (indirect), WMO
Community vulnerability indicators (heat discomfort, water access issues)	National portals or social-science repositories	Zenodo	Climate Explorer, CDS (contextual)

4.3 Mission: Cancer

In the Cancer Mission, citizen science plays a more limited but still valuable role compared with other Missions. Strict regulatory, ethical, and privacy requirements mean that most cancer-related data, such as clinical imaging, pathology results, or genomic files, must be generated and processed within formal medical and research settings. Citizen scientists cannot directly contribute to the major cancer repositories identified in D4.1 (TCIA, GDC, EGA) which operate under controlled access.

However, citizen science can contribute complementary, non-clinical datasets that provide real-world insights not captured in biomedical systems. These include patient-reported outcomes (symptoms, side effects, daily functioning), patient-generated health data from apps or wearables, lifestyle and behavioural factors relevant for cancer prevention, environmental exposure observations (radon, UV, air quality), and community-level insights on screening, awareness, and barriers to care. These datasets can support prevention, improve understanding of treatment impacts, and offer more patient-centred perspectives, if collected, anonymised, and published responsibly.

From D4.1, the key cancer-relevant repositories are:

- The Cancer Imaging Archive
- PatientsLikeMe
- Genomic Data Commons (GDC)
- European Genome-phenome Archive (EGA)

Only PatientsLikeMe directly accepts citizen science data. All other repositories rely on medical governance and controlled-access procedures, reinforcing the need for careful anonymisation, clear consent, and appropriate publishing routes.

4.3.1 Mission-specific data and interoperability requirements

Because cancer-related data often involve personal or sensitive information, data interoperability for this Mission focuses not only on technical standards but also on ethical, legal, and privacy safeguards (Table 5). Projects should meet the following cross-cutting requirements:

- **Strict anonymisation and pseudonymisation**, removing direct and indirect identifiers (e.g., precise address, free-text notes that could reveal identity).
- **Explicit informed consent**, including clear information on data reuse, confidentiality, and withdrawal.
- **Coarse spatial resolution** when mapping exposures or lifestyle factors (e.g., municipality or grid cells rather than exact coordinates).
- **Use of patient-outcome vocabularies**, such as PRO-CTCAE, PROMIS, or other recognised symptom/quality-of-life instruments.
- **Consistent categorisation** for behavioural and lifestyle data (activity, diet, smoking status, screening participation).
- **Clear metadata** describing the population (age bands, gender categories), methods of data collection, and potential limitations.
- **Structured, machine-readable formats**, such as CSV or JSON, avoiding spreadsheets with embedded free text.
- **Open licences** wherever ethically appropriate; if sensitive, use controlled sharing conditions (e.g., metadata open, data restricted).

Table 5: Key citizen science data types and interoperability requirements for the Cancer Mission

Type of Cancer-Relevant CS Data	Essential Metadata & Attributes	Relevant Standards / Conventions	How Mission Actors Use It
Patient-reported outcomes (symptoms, treatment side effects, fatigue, wellbeing)	Age band, gender, symptom type, severity, timestamp, collection method	PRO-CTCAE; PROMIS; PPSR-Core metadata	Supports quality-of-life studies, treatment tolerance, survivorship
Patient-generated health data (wearables, sleep, heart rate, activity)	Device type, sampling frequency, timestamps, anonymised ID	OGC SensorThings (optional);	Supports personalised care and long-term monitoring
Lifestyle & behavioural data (diet, activity, smoking status)	Categories, timestamp, age band, data collection method	WHO behavioural classifications; Dublin Core for high-level metadata	Supports prevention, behavioural risk factor analysis
Environmental exposures (radon, UV, local air quality)	Coarse location, timestamp, measurement method or proxy	INSPIRE environmental parameters; ISO 19115 spatial metadata	Helps identify exposure patterns linked to cancer risks
Screening behaviour & community insights (barriers to screening, perceptions)	Age band, gender category, timestamp, anonymised respondent ID	PPSR-Core for dataset metadata	Supports targeted cancer screening interventions
Lived-experience narratives (diagnosis stories, care experience)	Anonymised transcripts, thematic codes, timestamp (optional)	Qualitative research standards; Dublin Core	Supports patient-centred research and Mission engagement

4.3.2 Repository options for Citizen Science Data Relevant to the Cancer Mission

Because most cancer-specific repositories do not accept direct public submissions, citizen science data must follow indirect, ethically appropriate pathways. Table 6 outlines realistic publication options and how they align with the repositories identified in D4.1.

Table 6: Repository options for citizen science datasets supporting the Cancer Mission

Type of Cancer-Relevant CS Data	Preferred Publication Route	Alternative Route	Which Cancer Repositories Can Reuse / Compare It (Indirectly)
Patient-reported outcomes (symptoms,	PatientsLikeMe (direct), or national public-health portals	Zenodo (aggregated &	PatientsLikeMe (direct); indirect reuse by researchers using

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Type of Cancer-Relevant CS Data	Preferred Publication Route	Alternative Route	Which Cancer Repositories Can Reuse / Compare It (Indirectly)
wellbeing, fatigue, treatment side effects)		anonymised only)	GDC/EGA metadata context
Patient-generated health data (wearables, sleep, heart rate, physical activity)	PatientsLikeMe (if compatible) or secure national health data portals (metadata only)	Zenodo (aggregated, anonymised, non-identifiable time series)	Indirect contextual reuse by research groups working with GDC/EGA
Lifestyle & behavioural data (diet, smoking, activity)	National public health data portals, behavioural science repositories	Zenodo	Indirect contextual reuse by prevention researchers (GDC/EGA context)
Environmental exposures (radon, UV, local air quality)	National environmental portals or thematic exposure databases	Zenodo or GKH (when EO-linked)	Indirect reuse by risk-factor studies alongside environmental layers
Screening behaviour & community insights	Social-science repositories (e.g., Qualitative/Open Data Service), national public health portals	Zenodo	Indirect reuse by population-screening programmes
Lived-experience narratives	Qualitative data repositories (restricted access if needed)	Zenodo (only if anonymised & ethically cleared)	PatientsLikeMe (narrative integration); qualitative research

4.4 Mission: Climate-Neutral and Smart Cities

The Climate-Neutral and Smart Cities Mission depends on fine-grained, local data on mobility, emissions, public-space quality, energy use, environmental conditions, accessibility, and citizen behaviour. Citizen science plays a significant role in this mission because urban transformation requires detailed, real-time or high-resolution observations that residents can generate at scale. Typical contributions include air quality and noise measurements, mobility and traffic mapping, energy use patterns, urban heat and micro-climate data, public-space mapping, photographic and visual evidence, and community-led audits of accessibility, walkability, safety, and liveability.

D4.1 identified several repositories central to this mission's data ecosystem:

- Smart Cities Marketplace
- Smart Citizen Platform
- Zenodo (Urban Projects)

- OpenStreetMap (OSM)
- KM4City Ecosystem
- Panoramax

These repositories vary in their openness and data types. OSM and Panoramax directly accept citizen science data, supporting collaborative mapping and geospatial imagery. Zenodo is widely used by EU urban initiatives for structured dataset publication. Smart City Platform, Smart Cities Marketplace, and KM4City do not accept raw uploads but reuse, harvest, or integrate cleaned, documented datasets through project partners, APIs, or federated tools. This means citizen science datasets must be open, georeferenced, well-structured, and documented to support integration into mission-aligned urban digital ecosystems.

4.4.1 Mission-specific data and interoperability requirements

To ensure smart-city datasets generated by citizen science (Table 7) can be integrated into mobility planning tools, digital twins, sensor networks, and municipal analytics systems, projects supporting this Mission should meet the following cross-cutting requirements:

- **Street-level geolocation** using WGS84, including spatial accuracy and capture method (GPS, imagery, survey).
- **ISO 8601 timestamps** and measurement intervals for environmental and mobility sensors.
- Use of **urban vocabularies and ontologies**, including:
  - **OSM tagging schema** for features such as crossings, kerbs, cycle lanes, barriers, trees, benches, lighting, and pavements.
  - **ISO 37120 / 37122** indicators where relevant (urban services and smart city metrics).
- **Sensor metadata** for air quality, noise, temperature, mobility, or environmental sensors (device model, sampling frequency, calibration notes).
- **Clear accessibility descriptors**, such as slope, obstruction type, surface quality, and gradient.
- **Image metadata** (EXIF: geotag, timestamp, orientation) for street-level photography used in mapping or digital twins.
- **Machine-readable formats** including CSV, JSON, GeoJSON, and MBTiles (if providing tiled map outputs).
- **Licensing** compatible with reuse across smart-city platforms:
  - **ODbL** for OSM edits
  - **CC BY or CC0** for open datasets on Zenodo or national portals

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table 7: Key citizen science data types and interoperability requirements for the Climate-Neutral and Smart Cities Mission

Type of Urban CS Data	Essential Metadata & Attributes	Relevant Standards / Conventions	How Mission Actors Use It
Urban environmental measurements (air quality, noise, heat)	Coordinates, timestamp, units, sensor model, frequency	OGC SensorThings; INSPIRE environmental parameters	Feeds emissions inventories, heat maps, noise models
Mobility observations (walkability, cycling, accessibility)	Coordinates, route traces, timestamps, barrier types	OSM tags, ISO 37120	Supports active mobility planning and accessibility audits
Public-space features (trees, benches, lighting, crossings)	Point/line geometry, feature type, attributes (e.g., height, slope)	OSM tagging schema	Supports digital twins and public-space management
Urban imagery (street photos documenting heat islands, obstructions, shading)	EXIF metadata, timestamp, location, thematic labels	EXIF; OSM-compatible annotations	Enables computer vision and urban feature extraction
Community perceptions (liveability, safety, comfort)	Timestamp, geotag (optional), category labels, anonymised ID	PPSR-Core for metadata	Supports participatory planning and social sentiment mapping
Energy-related behaviour (home heating/cooling practices)	Time, category, anonymised ID, coarse location	ISO 37120 domains	Supports behaviour modelling in local energy transition

4.4.2 Repository options for Citizen Science Data Relevant to the Cities Mission

Citizen science datasets supporting the Climate-Neutral and Smart Cities Mission are typically integrated into urban data ecosystems through indirect reuse rather than direct ingestion into mission platforms. The repository options presented below outline (Table 8) the most appropriate publication routes for interoperable citizen science data, enabling reuse by smart-city infrastructures, urban analytics platforms, and mission-aligned initiatives identified in D4.1.

Table 8: Repository options for citizen science datasets supporting the Climate-Neutral and Smart Cities Mission

Type of Urban CS Data	Preferred Publication Route	Alternative Route	Which Smart-City Repositories Can Reuse / Compare It
Urban environmental measurements (air quality, noise, heat)	City or national environmental open data portals; Smart Citizen Platform (if compatible)	Zenodo	Smart Citizen Platform, KM4City
Mobility observations (walkability, cycling, accessibility)	OpenStreetMap (OSM) for mapped features; city mobility portals for datasets	Zenodo	OSM (direct), KM4City
Public-space features (trees, benches, lighting, crossings)	OSM (direct edits using OSM tagging schema)	-	OSM, KM4City
Urban imagery (street photos documenting shading, obstructions, heat islands)	City open data portals (if available)	Zenodo (image collections with EXIF metadata)	Smart Citizen Platform (indirect), KM4City
Community perceptions (liveability, safety, comfort)	Social-science repositories or city participation platforms	Zenodo	Smart Cities Marketplace
Energy-related behaviour (home heating/cooling practices)	City or national energy data portals	Zenodo	Smart Cities Marketplace

4.5 Mission: A Soil Deal for Europe

The Soil Deal Mission requires detailed, harmonised evidence on soil health across Europe, including soil structure, organic carbon, contamination, biodiversity, erosion, moisture, and land management practices. Citizen science contributes directly to the Soil Deal Mission by enabling widespread soil observations that are otherwise costly or impractical through traditional monitoring. Projects measuring soil moisture, pH, organic carbon, biodiversity, structure, contamination, and land management practices help build detailed, localised soil datasets. Citizen science data can support restoration planning, early detection of degradation, and improved soil literacy, all of which are key Mission objectives.

From D4.1, the key Soil deal for Europe-relevant repositories are:

- European Soil Data Centre (ESDAC)
- ISRIC-World Soil Information
- LUCAS Soil

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

- OpenLandMap

These repositories do **not** accept direct public submissions. They reuse citizen science data only when observations are published openly using clear geospatial metadata, harmonised soil descriptors, and transparent sampling methods.

4.5.1 Mission-specific data and interoperability requirements

To ensure that citizen science-based soil observations (Table 9) can be combined with harmonised soil data models and reused alongside institutional measurements, projects supporting this Mission should meet the following cross-cutting requirements:

- **Precise geolocation** (WGS84) with INSPIRE-aligned metadata for sampling method, depth, accuracy and land cover.
- **Sampling metadata**, including time, depth, soil condition at sampling, and simple site descriptors (e.g., slope, aspect).
- **Standard soil descriptors** such as texture (FAO/USDA), colour (Munsell or simplified categories), structure, compaction and organic material presence.
- **Biological indicators** (e.g., earthworms, macrofauna, root density) encoded as counts or categories.
- **Erosion indicators** (sheet, rill, gully; severity categories).
- **Sensor metadata** for soil moisture or temperature measurements (model, calibration, frequency).
- **Open formats** (CSV, JSON, GeoJSON) and clear variable names.
- **Transparent documentation** of protocols, validation steps, and limitations.
- **Open licences** (CC BY or CC0) unless agricultural privacy concerns apply.

Table 9: Key citizen science data types and interoperability requirements for the Soil Deal for Europe Mission

Type of Soil-Relevant CS Data	Essential Metadata & Attributes	Relevant Standards / Conventions	How Mission Actors Use It
Basic soil condition measures (texture, colour, structure, compaction)	Sampling depth, method, land cover, horizon (if known)	FAO/USDA texture classes; Munsell (optional)	Complements LUCAS and ESDAC soil health indicators
Soil moisture & temperature (sensor or manual)	Coordinates, timestamp, units, sensor model, calibration details	INSPIRE environmental parameters; OGC SensorThings	Supports drought, irrigation, and soil hydrology analysis
Soil biodiversity proxies (earthworms, macrofauna, root density)	Depth, method (e.g., spade test), counts or categories	FAO soil health biological indicators	Strengthens soil health and biological activity assessments

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Type of Soil-Relevant CS Data	Essential Metadata & Attributes	Relevant Standards / Conventions	How Mission Actors Use It
Organic matter proxies (colour, smell, litter depth)	Colour categories, cover type, depth	Categorical Munsell schemes	Supports understanding of organic carbon condition
Erosion observations (rills, gullies, bare soil)	Coordinates, timestamp, erosion type, severity	INSPIRE land cover/use; erosion typologies	Supports erosion-risk mapping and land degradation assessment
Land management practices (grazing, tillage, mulching)	Management type, frequency, season	FAO/CAP land-use categories	Provides essential context for interpreting soil health
Simple functional tests (infiltration, slaking)	Test method, volume/time, slope	Local soil monitoring protocols	Supports functional soil health indicators

4.5.2 Repository options for Citizen Science Data Relevant to the Soil Mission

Citizen science datasets relevant to the Soil Deal for Europe Mission are typically not ingested directly into central European soil repositories but can nevertheless support these infrastructures when published openly with appropriate metadata and documentation. The repository options outlined (Table 10) below reflect the pathways through which interoperable citizen science soil data can be discovered, compared, and reused alongside institutional soil datasets, as identified in D4.1.

Table 10: Repository options for citizen science datasets supporting the Soil Deal for Europe Mission

Soil CS Data Category	Preferred Route	Alternative Route	Which Soil Repositories Can Reuse It (Indirectly)
Soil condition (texture, colour, structure)	National agri/environment portals	Zenodo or GKH	ESDAC, OpenLandMap
Soil moisture & temperature	National hydrology/agri portals	Zenodo or GKH	OpenLandMap
Soil biodiversity	Institutional soil platforms	Zenodo	ESDAC (contextual)
Erosion observations	National environmental portals	Zenodo or GKH	OpenLandMap, ESDAC
Land management	CAP/agri portals	Zenodo	ESDAC

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Soil CS Data Category	Preferred Route	Alternative Route	Which Soil Repositories Can Reuse It (Indirectly)
Functional soil tests	Regional soil monitoring portals	Zenodo or GKH	ESDAC

4.6 Mission: Restore Our Ocean and Waters

The Oceans & Waters Mission depends on dense, multi-scale environmental observations that are often costly or difficult to obtain through institutional monitoring alone. Citizen science contributes across this spectrum: measuring freshwater quality, documenting coastal erosion and pollution, observing marine biodiversity, reporting algal blooms and fish kills, and capturing hydrological conditions in small streams and lakes. These datasets can support the Mission's goals of reducing pollution, restoring ecosystems, and enabling nature-based solutions, when they follow recognised oceanographic, hydrological, and biodiversity standards.

From D4.1, the key Restore our oceans and waters-relevant repositories are:

- European Marine Observation and Data Network (EMODnet)
- SeaDataNet
- iNaturalist (Marine)
- Copernicus Marine Service (CMEMS)
- Ocean Biodiversity Information System (OBIS)
- PANGAEA
- FreshWater Watch
- Global Lake Ecological Observatory Network (GLEON)
- Global Runoff Data Centre (GRDC)
- HydroShare
- AquaSTAT

4.6.1 Mission-specific data and interoperability requirements

Citizen science-based water and marine data (Table 11) must align with established hydrological, oceanographic, and biodiversity standards.

- **WGS84 geolocation** (with depth where relevant)
- **ISO 8601 timestamps**
- **Standard marine/hydrology units** (mg/L, NTU, °C, PSU, m/s)
- **Darwin Core** for marine species observations
- **OGC SensorThings or OGC SWE** for sensor data
- **Method and calibration metadata**
- **Open formats** (CSV, JSON, GeoJSON; **NetCDF** for gridded/modeled datasets)
- **Clear documentation of sampling method**
- **Open licence** (CC BY / CC0)

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table 11: Key citizen science data types and interoperability requirements for the Restore Our Ocean and Waters Mission

Type of Water-Relevant CS Data	Essential Metadata & Attributes	Relevant Standards	How Mission Actors Use It
Freshwater quality (nitrate, turbidity, clarity, pH)	Coordinates, timestamp, units, sensor model/calibration	OGC SensorThings; INSPIRE	Supports lake & river monitoring, complements official datasets
Marine biodiversity (species observations)	Species name, date, location, method	Darwin Core	Feeds OBIS and marine biodiversity indicators
Coastal impacts (erosion, storm surge, litter)	Location, timestamp, category, photos	INSPIRE land/sea interface; EXIF	Supports coastal erosion and MSFD indicators
Hydrological observations (flow, water level)	Coordinates, depth, timestamp, method	ISO 19115; OGC SWE	Supports basin modelling and flood risk
Oceanographic proxies (sea surface temp, transparency)	Timestamp, coordinates, units, sensor metadata	OGC SensorThings; CF conventions	Supports CMEMS model validation
Community observations (algal blooms, fish kills)	Location, timestamp, photos, category	EXIF; PPSR-Core metadata	Adds early-warning signals for water agencies

4.6.2 Repository options for Citizen Science Data Relevant to the Restore our Oceans and Waters Mission

The Restore Our Ocean and Waters Mission benefits from a diverse ecosystem of marine, freshwater, and hydrological data infrastructures, many of which already integrate citizen science contributions either directly or through secondary reuse pathways. The repository options presented below (Table 12) illustrate how interoperable citizen science datasets can be published and subsequently reused by mission-aligned marine and water data systems identified in D4.1.

Table 12: Repository options for citizen science datasets supporting the Restore Our Ocean and Waters Mission

Water CS Data Type	Preferred Route	Alternative Route	Repositories Using It Indirectly
Freshwater quality	FreshWaterWatch, GLEON	Zenodo or GKH	EMODnet (context)
Marine biodiversity	iNaturalist → OBIS	Zenodo	OBIS

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Water CS Data Type	Preferred Route	Alternative Route	Repositories Using It Indirectly
Coastal impacts	National marine/litter portals	Zenodo	EMODnet, CMEMS
Hydrological observations	National hydrology portals; HydroShare	Zenodo or GKH	GRDC
Oceanographic proxies	National marine portals	Zenodo or GKH	CMEMS
Community observations	National environmental portals	Zenodo	EMODnet (context)

Citizen science meaningfully supports the Oceans & Waters Mission by generating granular, local water and marine observations. Structured using Darwin Core, INSPIRE, and OGC standards and published via domain or national portals, these datasets can inform pollution reduction, biodiversity restoration, and basin-scale modelling.

5 Guidelines for citizen science data interoperability

5.1 Purpose of the Guidelines

This chapter provides a concise set of practical guidelines that citizen science projects can follow to ensure their data are interoperable, reusable, and ready for integration into the data ecosystems of the five EU Missions. The guidelines directly address recurring data-related barriers identified in D2.1 and D2.2, particularly fragmented metadata practices, limited standards adoption, and unclear pathways for data reuse beyond individual projects. The guidance presented below synthesizes the recurring metadata, quality, and publication requirements observed across Missions, offering a simple and universal reference that complements the mission-specific pathways presented in Chapter 4.

5.2 Guidelines and Interoperability Checklist

The following combined guidelines (Table 13) and checklist offer a single, practical reference that citizen science projects of any type can apply. They reflect the shared interoperability needs identified across all five Missions and the repository requirements reviewed in D4.1. Projects that follow these steps will be better positioned to ensure that their data can be discovered, interpreted, combined, and reused by research communities, public authorities, and mission actors.

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table 13: Interoperability guidelines and checklist for citizen science projects

	Guidelines	Checklist
Data Format and Structure	Use open, machine-readable formats such as CSV, JSON, GeoJSON, and NetCDF for gridded outputs. Avoid proprietary formats or spreadsheets with formulas, merged cells, or embedded objects. Keep variable names, units, and data types consistent across the dataset. Use categorical fields or structured inputs where possible; limit free-text only to contextual comments.	☐ Data provided in CSV/JSON/GeoJSON (or NetCDF where relevant) ☐ Variables use consistent names, units, data types ☐ No core variables in free text
Standardised Metadata	Include clear descriptive metadata: dataset title, overview, purpose, and provenance. Provide methodological metadata: sampling protocol, instructions given to volunteers, instrument or sensor details, calibration procedures, and known limitations. Always include WGS84 coordinates with spatial accuracy and method of capture. Use ISO 8601 timestamps, sampling intervals, and overall temporal coverage. Prepare a data dictionary that describes every variable, allowed values, units, and definitions. Add versioning and document changes to the data collection process.	☐ Descriptive, methodological, spatial, temporal, and structural metadata included ☐ Coordinates in WGS84; timestamps in ISO 8601 ☐ Data dictionary provided ☐ Version history documented
Standards, Vocabularies, and Identifiers	Use relevant domain standards where they apply: PPSR-Core, INSPIRE, ISO 19115, Darwin Core, OSM tagging schema, PRO-CTCAE, PROMIS (where health-related), etc. Apply controlled vocabularies and classification schemes rather than ad hoc labels. Use persistent identifiers for sites, sensors, observation series, or anonymised contributors as appropriate.	☐ Domain standards applied consistently ☐ Controlled vocabularies used for categorical fields ☐ Persistent IDs assigned (DOI, site ID, sensor ID, or equivalent)

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Quality and Processing	Conduct basic validation and quality checks during and after data collection. Document any processing, filtering, cleaning, or transformations applied to the data. Provide uncertainty indicators or quality flags when applicable. Preserve raw data separately from cleaned or standardised versions. Maintain version control for updated or expanded datasets.	☐ Quality checks and validation documented ☐ Uncertainty flags included ☐ Raw and cleaned data stored separately ☐ Version control maintained
Publication and Licensing	Publish data through the most suitable repository route: - Domain repositories that accept citizen submissions (e.g. GBIF, OpenStreetMap, FreshWaterWatch, GLEON, HydroShare). - National or thematic portals for environmental, hydrology, marine, soil, or urban data. - Zenodo or GKH when domain or national routes are unavailable or when a DOI is required. Apply an open licence (CC BY or CC0) to allow broad reuse. Ensure that repository entries include complete metadata and links to project websites, dashboards, or protocols.	☐ Open licence applied (CC BY or CC0) ☐ Repository selected based on data type and mission relevance ☐ Metadata complete upon upload ☐ DOI or persistent identifier assigned

Following these guidelines ensures that citizen science datasets are clear, structured, well-documented, and ready for reuse across diverse European data ecosystems. Consistent use of open formats, complete metadata, recognised standards, transparent methods, and appropriate repositories significantly increases the scientific value and policy relevance of citizen science data and strengthens their contribution to the EU Missions.

6 Conclusion

Citizen science has the capacity to generate locally relevant, high-resolution and socially meaningful data, but its contribution to European research and policy depends on the extent to which these datasets can be integrated into existing infrastructures. This deliverable demonstrates that interoperability is not a single technical requirement, but a coordinated effort across planning, data management, documentation, ethics, licensing and repository selection. The guidelines and checklist provided here offer a practical pathway to ensure that citizen science data can be reliably reused, interpreted and combined with institutional sources.

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

Insights from D4.1 confirm that Europe hosts a mature and diverse repository landscape, yet many platforms lack clear routes for citizen data ingestion. The guidance developed in D4.2 enables projects to bridge this gap by aligning their workflows with established standards and repositories, increasing mission-readiness and long-term sustainability. By adopting interoperable formats, standardised metadata, transparent licences and well-documented validation processes, citizen science initiatives enhance both scientific credibility and policy relevance.

The guidance presented in this deliverable is intended to support citizen science projects that wish to scale their initiatives, either by expanding in size or reach, or by enabling reuse of their data for policymaking, research, and other broader applications. At the same time, it is recognised that not all citizen science initiatives seek or benefit from scaling or interoperability. Many projects are intentionally local or community-centred and may prioritise contextual relevance, local ownership, or sensitivities around data sharing. These choices are valid, and the guidance provided here is not intended to impose interoperability where it is not desired, but to support projects that choose to pursue it.

Several limitations should be acknowledged. This deliverable focuses primarily on technical and procedural aspects of interoperability and cannot fully address the social, institutional, and governance factors that also shape data practices in citizen science. In addition, standards, repositories, and tools continue to evolve, meaning that this guidance reflects current practice rather than a fixed endpoint. Looking ahead, emerging technologies such as artificial intelligence (AI) may offer opportunities to support interoperability, for example through metadata generation, data harmonisation, or quality checks. These opportunities are accompanied by challenges related to transparency, bias, dependency on proprietary systems, and trust. Future work should therefore explore how AI can be applied responsibly and transparently, in ways that align with the values of citizen science and complement, rather than replace, human and community oversight.

As CROPS advances towards its wider objective of supporting upscaling across the ERA, these guidelines provide a foundation for consistent data practices that can be shared, adapted and propagated across communities. They also inform subsequent work on interoperability, funding, and sustainability within WP4, and reinforce the role of citizen science as a meaningful contributor to the Horizon Europe Missions. The result is a more connected and accessible data ecosystem, enabling citizen science evidence to inform decisions, stimulate innovation and underpin collective action across Europe.

Funded by the European Union

Project funded by the European Union (GA 101131696). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.