Deliverable 4.1

An analysis of open data repositories

30.06.2025

Kaushiki Das and James Sprinks
Earthwatch Europe

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

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.

Document history

Version	Date	Comment	Modification made by
01	30/05/2025	Initial draft	Earthwatch (K. Das)
02	25/06/2025	Minor revisions	Earthwatch (J. Sprinks)
03	30/06/2025	Final revisions	Earthwatch (J. Sprinks)

To cite this document:

Das, K., Sprinks, J. (2025). D4.1: An analysis of open data repositories. 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.

Table of Contents

Executive Summary **6**
1. Introduction **8**
1.1. Linkages across the project 9
2. Methodology **10**
2.1. Repository Selection Criteria 11
2.2. Comparative Framework 12
3. Review of open data repositories **13**
4. Review of open data repositories **18**
4.1. Adaptation to climate change 18
4.2. Cancer 21
4.3. Restore our Ocean and Waters by 2030 23
4.4. 100 Climate-neutral and smart cities by 2030 26
4.5. A Soil Deal for Europe 28
5. CROPS and ScienceUs project practices **32**
5.1. The 25 Projects Identified: Purpose, Progress, and Positioning 32
5.1.1. Type of data shared 32
5.1.2. Openness and accessibility 33
6. Social media for citizen science data dissemination **36**
7. Comparative Analysis **39**
7.1. Aim 39
7.2. Method 40
7.3. Results 41
8. Best practices for open data repositories **46**
9. Limitations **48**
10. Recommendations **49**
11. Extending the framework: cross-cutting analytical dimensions **51**
11.1. Metadata standardisation 51
11.2. Machine-Readability and Automation 51
11.3. Equity and Inclusion 52
11.4. Policy Integration 52
11.5. Governance and Ethical Transparency 52
12. Conclusion **53**

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.

13. References 53

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.

Tables and figures

Table 1: General purpose citizen science repositories 13
Table 2: Evaluating general purpose repositories 14
Table 3: Overview of open and mission-specific repositories 15
Table 4: Digital data repositories on climate data 18
Table 5: Digital data repositories on cancer 22
Table 6: Digital data repositories on freshwater and ocean water 24
Table 7: Digital data repositories on climate neutrality 26
Table 8: Digital repositories on soil health 30
Table 9: Mapping of ScienceUs projects based on data type and accessibility 33
Table 10: Repositories supporting citizen-generated local data collection 36
Table 11: Repository-level comparison across the five analytical dimensions 43

Figure 1: PRISMA-inspired flow of repository selection 17
Figure 2: Radial charts of mission scores across five core dimensions (comparative and per each mission) 42

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 report presents a comparative analysis of open data repositories that support the five EU Missions: 'Adaptation to Climate Change', 'Cancer', 'Restore our Oceans and Waters', 'Climate-Neutral and Smart Cities', and 'A Soil Deal for Europe'. It evaluates how these repositories align with FAIR principles (Findable, Accessible, Interoperable, and Reusable) and support citizen science participation.

Using a desk-based review, the study applied a structured framework assessing five core dimensions - repository openness, citizen data submission, interoperability, reuse licensing clarity, and institutional sustainability - alongside cross-cutting criteria such as metadata standards, machine-readability, policy integration, and ethical governance.

Findings show that while repositories like EMODnet, Copernicus, and ESDAC perform strongly in technical and policy alignment, most platforms lack clear citizen submission channels, consistent reuse terms, or visible governance frameworks. Participatory platforms such as iNaturalist and OpenStreetMap highlight best practices but are not yet mainstreamed across Missions.

The report identifies best practices and recommends improving cross-repository interoperability, expanding participatory infrastructure, clarifying licensing, and ensuring long-term sustainability. These steps are essential to transform open repositories into inclusive, ethical, and durable infrastructure for citizen science and mission delivery.

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

Open data repositories are digital platforms that store and disseminate datasets from scientific projects, including citizen science initiatives. They play a crucial role in ensuring transparency, accessibility, and collaboration in science by making volunteer-collected data publicly available. Open data is widely recognised as a catalyst for innovation and discovery, serving as "fuel for innovation and scientific discovery" while also promoting a more transparent and collaborative research environment. In practice, openly available data have enhanced scientific understanding and informed policy decisions – for example, shared data on environmental and health parameters enable more informed decision-making and targeted interventions by authorities. This is particularly evident in fields like environmental monitoring, public health, and urban planning, where access to diverse datasets allows researchers and policymakers to integrate knowledge and respond effectively to complex issues. Equally important is the role of repositories in the long-term preservation, validation, and reuse of citizen science data. Dedicated open data repositories provide a secure home for datasets, ensuring they remain preserved, accessible, and citable over time. By adhering to curation standards, repositories help validate data quality and enable other scientists to reproduce analyses or combine data across studies. One major feature of open datasets is that it allows others to scrutinise and build upon the results, which maximises the collective scientific benefit of citizen contributions. Such platforms thus uphold core open science principles: they facilitate reproducibility, transparency, and reuse by allowing anyone to verify results or repurpose data in novel research.

In the year 2021, the European Union formalised five missions to achieve by 2030, each addressing a major societal challenge. These five missions are:

1. **'Adaptation to Climate Change'** – supporting at least 150 European regions and communities to become climate resilient by 2030.
2. **'Cancer'** – working with Europe's Beating Cancer Plan to improve the lives of more than 3 million people by 2030 through better prevention, treatment, and solutions for longer, healthier lives.
3. **'Restore Our Ocean and Waters'** – aiming to protect and restore aquatic ecosystems and ensure sustainable management of oceans, seas, and waters by 2030.
4. **'Climate-Neutral and Smart Cities'** – advancing 100 European cities towards climate neutrality by 2030, fostering smart, sustainable urban development.
5. **'A Soil Deal for Europe'** – establishing 100 living labs and lighthouses to lead the transition toward healthy soils by 2030, promoting sustainable land management practices.

Each mission actively seeks to engage citizens in its implementation. Thus, citizen science is a key vehicle- contributing to the missions' objectives by facilitating grassroots data collection and public participation. This participation allows citizens to collect localised data and contribute to online data repositories. Through participatory efforts,

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.

citizen science provides the EU Missions with extensive real-time data and grassroots innovation, while also empowering communities to take part in reaching the missions' targets.

Given the importance of open data in both citizen science and the EU Missions, **the purpose of this deliverable is to review and analyse open data repositories that are relevant to the five EU Missions. The report focuses on repositories holding citizen science data and examines how these platforms support data accessibility**, what their data submission and curation requirements are, and what policies they have for data reuse and attribution. By evaluating these aspects, we aim to assess how well current repositories facilitate open citizen science data sharing in alignment with FAIR principles – ensuring data are Findable, Accessible, Interoperable, and Reusable. Identifying strengths and gaps in repository practices will allow us to recommend improvements to enhance data interoperability across platforms and mission domains. Ultimately, the findings of the report will provide guidance on improving open data practices – for example, suggesting standard metadata or licensing policies to better enable data integration and reuse.

1.1. Linkages across the project

This deliverable is part of **work package 4: orchestration** – maximising uptake and sustainability when upscaling citizen science. It contributes specifically to task 4.1: Identification and evaluation of open data repositories and infrastructures. The objective of this task is to assess existing open data repositories relevant to citizen science initiatives, particularly in the context of the five EU Missions. The analysis focuses on repository features such as the types of data shared, submission requirements, openness, reuse policies, and alignment with FAIR principles. By conducting this assessment, the deliverable provides foundational insights into the data infrastructure landscape that will support the long-term sustainability and impact of up scaled citizen science activities across Europe.

In the wider context of the CROPS project, which aims to curate, replicate, orchestrate, and propagate citizen science activities at a transnational level, this deliverable contributes to the orchestration strand by identifying practical barriers and enablers related to data sharing, interoperability, and accessibility. Its findings will inform Task 4.2 and the associated D4.2: Guidelines for citizen science data interoperability (M24), which focuses on developing protocols and guidance for the ethical and inclusive use of open data infrastructures in citizen science, as well as the wider replication and propagation efforts across WPs 3 and 5.

The structure of this deliverable follows a logical progression. It begins with an executive summary and a methodology section outlining the research approach, followed by a comprehensive review of general-purpose and mission-specific repositories. It then presents a comparative analysis across key dimensions, identifies best practices and

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.

limitations, and concludes with a set of actionable recommendations. This structure is intended to provide both a strategic overview and detailed guidance for stakeholders involved in citizen science data management.

2. Methodology

This study adopts a desk-based research approach, grounded in an extensive review of open data repositories relevant to citizen science initiatives within the five EU Missions. The process combined academic literature, policy documents, and outputs from previous EU-funded projects such as CS-TRACK and COS4CLOUD, which provided key insights into the evolving landscape of citizen science data infrastructures. Supplementing these sources, targeted online searches were conducted to identify repositories spanning domains central to the five Missions—climate change adaptation, cancer research, marine ecosystems, urban sustainability, and soil health. Key platforms reviewed included data.europa.eu, Zenodo, Dryad, GBIF, and EMODnet, alongside institutional repositories managed by government agencies, research organisations, and citizen science networks. Due to the focus on the EU missions, this analysis centres on the scientific data produced by citizen science initiatives and its flow, rather than personal data relating to the participants of citizen science activities (save behavioural data linked to quality and validity). The consideration of personal data raises a number of considerations related to ethics, privacy and data protection law, the support of which is part of other tasks in the CROPS project. Of course, this distinction is not always easy to separate, especially when considering health data and the EU 'Cancer' mission, which will be discussed in later sections of this document.

The repository identification process was further supported by AI-assisted tools, specifically ChatGPT (OpenAI)[1] and Elicit AI[2], to enhance the breadth and efficiency of preliminary scoping. These tools were used to generate initial lists of potentially relevant open data repositories aligned with the FAIR principles (Findable, Accessible, Interoperable, and Reusable) and the thematic areas of the five EU Missions. Prompts included targeted queries such as *"Which open data repositories support citizen science in climate adaptation?"* and *"List repositories enabling citizen-contributed environmental or genomic data."* The AI-generated suggestions were not taken at face value but served as a starting point for manual verification. Each repository identified via these tools was cross-checked through official documentation and prior project outputs (e.g., COS4CLOUD, CS-TRACK) to assess suitability for inclusion. CROPS is dedicated to the transparent use of AI, and recognises the potential issues in its use; this approach allowed us to maintain research rigour through intensive human-led validation.

[1] https://openai.com/chatgpt/overview/
[2] https://elicit.com/

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.

2.1. Repository Selection Criteria

Repositories were selected using a defined set of criteria to ensure relevance to both citizen science and the objectives of the five EU Missions:

- **Relevance to mission domains:** Citizen Science data can play a supportive role in achieving the EU mission objectives, for instance observations can help implement the Water Framework Directive, and it has been found that 75% of invasive species projects align directly with EU regulatory processes (Price-Jones et al., 2022). Therefore, repositories had to either host datasets generated through citizen science or be widely used in fields directly linked to one or more of the Missions—namely 'Adaption to Climate Change', 'Cancer', 'Restore our Ocean and Waters', 'Climate-Neutral and Smart Cities', and 'A Soil Deal for Europe'.
- **Openness and accessibility:** Transparent data management and a willingness to share can build trust and enhance volunteer participation (Groom et al., 2017). Priority was given to repositories that provide free, public access to data, with transparent licensing policies—particularly those adhering to Creative Commons (CC0, CC BY) or Open Government Data (OGD) principles.
- **Citizen data submission and curation:** Repositories that explicitly allow or encourage data contributions from individuals, communities, or citizen science networks were emphasised, especially where clear curation and validation processes were in place.
- **Metadata quality and interoperability:** Controlled vocabularies, standardized metadata, and simplified licensing can clarify use rights and support data integration (Bowser et al., 2020). Preference was given to platforms using recognised metadata schemas such as **Darwin Core** (biodiversity), **ISO 19115** (environmental data), or **Dublin Core** (general datasets), and that support structured formats conducive to data integration.
- **Machine-readability and data exchange protocols:** End-to-end technical solutions, semantic resources, and open, machine-readable platforms can facilitate data discoverability, interoperability and reuse (Lush et al., 2024). Repositories offering machine-readable formats (e.g. CSV, JSON, NetCDF, GeoTIFF), API access, or compatibility with federated data-sharing protocols were considered particularly suitable.
- **Institutional sustainability:** Institutional sustainability and funding are critical for hosting open data repositories in citizen science. Open data reuse and collaborative repositories enable citizen science projects to have impact beyond their initial scope or funding cycle (Calyx, 2020). Repositories with persistent identifiers (e.g. DOIs), long-term funding, and reliable archiving mechanisms were favoured as they are more likely to ensure continued access and use of citizen science data beyond individual project timelines.

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.

2.2. Comparative Framework

To structure the analysis, a comparative framework was developed to assess repositories against key attributes that shape data usability, inclusivity, and long-term value. These attributes included the types of data hosted—ranging from climate indicators and biodiversity observations to genomic records, marine monitoring data, urban mobility datasets, and soil health metrics—and the submission requirements, particularly who is allowed to contribute data (e.g. individuals, institutions) and whether any technical or quality standards are applied. The framework also considered access and reuse policies, including whether data are openly accessible, subject to restrictions, or available via downloads, APIs, or real-time feeds.

Two further dimensions were emphasised: interoperability, which focused on the adoption of standard metadata formats, machine-readable structures, and compatibility with platforms such as INSPIRE and the European Open Science Cloud (EOSC); and institutional sustainability, which examined the presence of long-term funding, governance structures, and data preservation mechanisms.

In applying this framework, five core evaluation dimensions were defined:

- **Repository openness** refers to the extent to which datasets are publicly accessible without institutional restrictions, paywalls, or opaque access procedures.
- **Citizen data submission** captures whether and how repositories enable individuals, communities, or non-institutional actors to contribute data directly.
- **Interoperability** assesses the repository's use of standard metadata formats and data structures that support integration, exchange, and reuse across platforms.
- **Reuse licensing clarity** evaluates how clearly the repository communicates the legal terms under which data can be reused, including attribution and commercial use.
- **Institutional sustainability** reflects the presence of long-term funding, governance, and preservation mechanisms that ensure continued data access beyond individual project cycles.

Applying this structured framework enabled a comparative analysis of repositories across the five EU Missions, revealing both mission-specific strengths and shared challenges. The analysis offers a foundation for understanding how open data infrastructures currently support the ambitions of the EU Missions, and where further alignment with citizen science, open innovation, and policy integration is needed. Open data repositories play a foundational role in making datasets publicly accessible—promoting transparency in governance, accountability in research, and collaboration across sectors. By offering access to high-quality datasets, these platforms drive innovation and knowledge exchange, enabling businesses, researchers, and

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.

policymakers to develop evidence-based solutions. They also empower citizens and advocacy groups by enhancing participation and trust in science.

3. Review of open data repositories

This section presents a preliminary review of widely used general-purpose repositories that, while not tied to a specific EU Mission, provide open infrastructure suitable for citizen science contributions. These repositories serve as important reference points for assessing baseline practices in data openness, submission, and reuse.

Table 1: General purpose citizen science repositories

Name	URL	Best use case for citizen science
Dryad	https://datadryad.org/	Pollinator tracking, epidemiological studies
Figshare	https://figshare.com/	Noise pollution mapping, participatory urban research
Zenodo	https://zenodo.org/	Biodiversity monitoring, air pollution tracking
EOSC EU Node	https://open-science-cloud.ec.europa.eu/	Climate adaptation, land-use research, urban air quality monitoring

The following sections assess open data repositories based on four key criteria:

1. **Types of data shared:** Assesses the thematic scope and formats of stored datasets and their alignment with FAIR principles.
2. **Requirements for data submission:** Examines submission conditions, such as institutional affiliation, peer review, or fees, which affect inclusivity.
3. **Repository openness:** Reviews access conditions and whether data are openly available or restricted to specific user groups.
4. **Policy for data reuse:** Evaluates the clarity and permissiveness of licensing terms, especially the use of Creative Commons licences (e.g. CC0, CC-BY).

The four criteria used in this preliminary assessment—types of data shared, requirements for data submission, repository openness, and policy for data reuse—are grounded in the FAIR principles (Findable, Accessible, Interoperable, Reusable), as defined by Wilkinson et al. (2016), and have been widely operationalised in the evaluation of open data infrastructures. These dimensions are consistent with frameworks used by the European Open Science Cloud (EOSC) and RECODE project, which assess open access data practices in terms of accessibility, reuse licensing, and institutional conditions for participation (RECODE, 2014; EOSCpilot, 2019). Additionally, projects like COS4CLOUD and CS-TRACK have highlighted the importance of submission pathways and licensing clarity in supporting sustainable and inclusive citizen science 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.

Evaluating the types of data shared enables assessment of a repository's thematic breadth and relevance to diverse scientific and civic domains, while submission requirements directly influence the inclusivity of citizen engagement. Repository openness is a core tenet of FAIR accessibility, ensuring public users can retrieve data without barriers. Finally, reuse licensing determines the legal flexibility with which citizen-generated or publicly funded data can be shared, repurposed, or integrated into downstream research and policy tools. Collectively, these criteria offer a practical yet theory-aligned framework to assess the baseline openness and suitability of general-purpose repositories for citizen science engagement.

The following table provides a structured comparison of general-purpose repositories, evaluating them against these four criteria to highlight their suitability for Citizen Science initiatives:

Table 2: Evaluating general purpose repositories

Repository	Types of Data Shared	Requirements for Data Submission	Policy for Data Access	Policy for Data Reuse
Dryad	Numerical, geospatial, multimedia, software, sensor data	Open participation, assigns DOI, supports GitHub integration	Open access (CC0, CC-BY), no institutional restrictions	Flexible licensing (CC0, CC-BY), supports unrestricted reuse
Zenodo	Multimedia, datasets, figures, videos, presentations	Open participation, multiple file formats supported	Open access (CC0, CC-BY), public and restricted	Supports Creative Commons licensing for public
Figshare	Biological, ecological, health sciences data	Requires peer-reviewed publication, metadata standardisation	Open access (CC0), requires linked journal	CC0 license, allows unrestricted
EOSC EU Node	Federated EU research data across multiple domains	Varies by data source, some institutional	Varies by repository, promotes open access	Mixed licensing policies, depending on data

As we can see form the following table, Data types determine a repository's suitability for citizen science contributions, ranging from numerical and geospatial data to multimedia and sensor inputs. Submission requirements influence inclusivity, as some repositories impose restrictions such as peer-review validation, institutional affiliation, or submission fees, while others allow open participation. For instance, Dryad requires

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.

peer-reviewed links, whereas Zenodo and Figshare allow open submission. Access policies impact usability, with open-access platforms like Zenodo and Figshare promoting transparency, whereas some repositories impose restrictions due to ethical and privacy concerns, as seen in The Cancer Imaging Archive. Reuse policies affect data interoperability, with CC0/CC-BY licences supporting open collaboration, while restrictive licensing limits redistribution. Zenodo, Figshare, and the European Open Science Cloud EU Node facilitate broad data integration, whereas Dryad and The Cancer Imaging Archive impose constraints to ensure data quality and ethical compliance. These criteria help assess repository openness, accessibility, and sustainability, ensuring effective data sharing for citizen science initiatives. This section will attempt a literature review of the existing data repositories based on these four criteria's, categorised according to each EU Mission. Before moving on to the next section, table- presents a consolidated table of all open and mission-specific repositories reviewed in this deliverable.

Table 3: Overview of open and mission-specific repositories

Category	Repository name	Best use case	URL
Open data	Dryad	Pollinator tracking, epidemiological studies	https://datadryad.org/
Open data	Zenodo	Biodiversity monitoring, air pollution tracking	https://zenodo.org/
Open data	Figshare	Noise pollution mapping, participatory urban research	https://figshare.com/
Open data	EOSC EU Node	Climate adaptation, land-use research, urban air quality monitoring	https://open-science-cloud.ec.europa.eu/
Climate adaptation	Copernicus Climate Data Store (CDS)	Climate reanalysis, temperature trends, weather indicators	https://cds.climate.copernicus.eu/
Climate adaptation	Climate-ADAPT	Heatwaves, precipitation, water stress indicators	https://climate-adapt.eea.europa.eu/
Climate	DRMKC Risk	Floods, wildfires,	https://drmkc.jrc.ec.europa.eu/risk-

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.

adaptation	Data Hub	risk maps	data-hub
Climate adaptation	GBIF	Biodiversity and climate resilience data	https://www.gbif.org/
Climate adaptation	WMO Data Platform	Meteorological and hydrological data	https://community.wmo.int/activity-areas/wis/wis-2
Cancer	The Cancer Imaging Archive	Radiological images (MRI, CT, PET)	https://www.cancerimagingarchive.net/
Cancer	PatientsLike Me	Patient-reported outcomes, symptom tracking	https://www.patientslikeme.com/
Cancer	Genomic Data Commons (GDC)	Genomic and clinical data	https://gdc.cancer.gov/
Cancer	European Genome-phenome Archive (EGA)	DNA/RNA sequencing, phenotype data	https://ega-archive.org/
Ocean and waters	EMODnet	Marine	https://emodnet.eu/
Ocean and waters	Marine Data Archive (MDA)	General marine datasets	https://www.marine-data.eu/
Ocean and waters	SeaDataNet	Oceanographic and marine environmental data	https://www.seadatanet.org/
Ocean and waters	iNaturalist (Marine)	Marine species sightings by citizens	https://www.inaturalist.org/

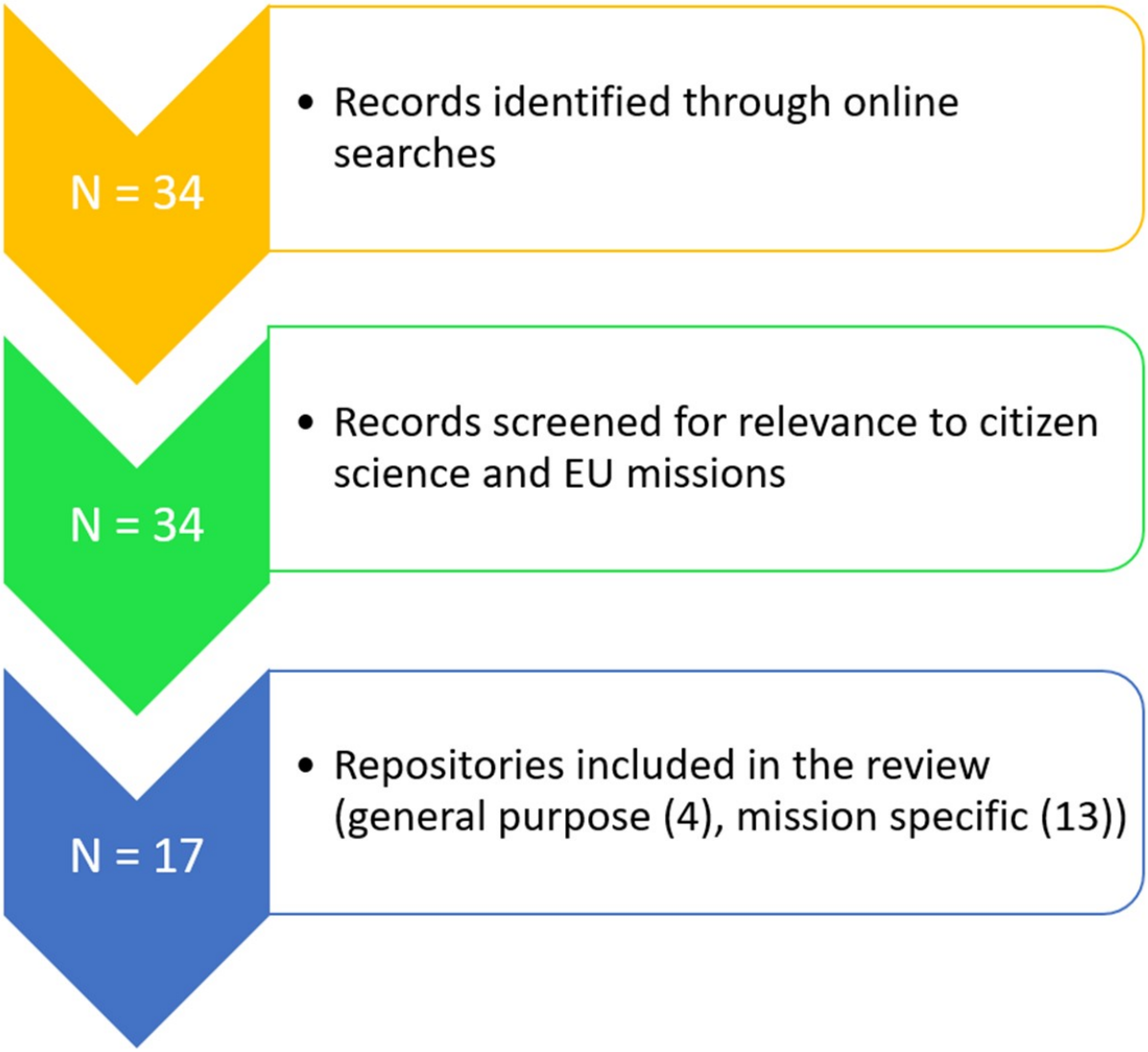
To ensure transparency in the repository selection process, we developed a structured identification and screening approach inspired by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework. A total of 34 repositories were initially identified through desk research, literature review, and project partner inputs. These included general-purpose open data platforms as well as repositories linked to each of the five EU Missions. All were screened for relevance to citizen science, openness, metadata standards, submission accessibility, and alignment with FAIR principles. Of these, 17 repositories met the inclusion criteria and were retained for in-

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.

depth analysis, comprising 4 general-purpose platforms and 13 mission-specific repositories. The flow of repository identification, screening, and inclusion is summarised in the PRISMA inspired diagram below.

Figure 1: PRISMA-inspired flow of repository selection



Of the 34 repositories initially identified through desk research and prior project outputs (e.g. COS4CLOUD, CS-TRACK), 17 met our inclusion criteria of openness, relevance to citizen science, and alignment with one or more of the five EU Missions. The remaining were excluded due to limited accessibility, lack of relevance, or insufficient documentation.

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.

4. Review of open data repositories

4.1. Adaptation to climate change

'The adaptation to climate change' mission under the EU Missions framework aims to help regions, cities, and communities prepare for and mitigate the adverse effects of climate change. Climate-related disasters such as extreme heat, droughts, floods, and wildfires are becoming more frequent and severe. This mission supports at least 150 European regions and communities in building resilience by 2030 through initiatives such as developing early warning systems, implementing nature-based solutions, and enhancing climate risk assessments. Achieving these objectives requires accurate and open climate data that can guide policymakers, researchers, and local communities in their decision-making. Without accessible data, stakeholders may struggle to understand climate risks, model future scenarios, or develop effective adaptation strategies. Open data repositories play a crucial role in bridging knowledge gaps, ensuring that relevant climate information is available to those who need it most.

Importance of open data repositories for climate adaptation efforts

Open data repositories are essential for climate adaptation efforts as they provide unrestricted access to high-quality climate datasets. These repositories support real-time monitoring, historical analysis, and predictive modelling—all crucial for preparing communities for climate-related challenges. Additionally, they foster collaboration between governments, scientists, and citizen scientists, ensuring that adaptation measures are data-driven and inclusive. Beyond facilitating research and policy decisions, open data repositories also promote transparency. By making climate data freely available, they enable local governments, NGOs, and businesses to develop targeted climate adaptation strategies without barriers. Moreover, compliance with FAIR principles ensures that datasets can be easily discovered and integrated across different adaptation initiatives.

Table 4: Digital data repositories on climate data

Repository	Types of Data Shared	Requirements for Data Submission	Policy for Data Access	Policy for Data Reuse
Copernicus Climate Data Store (CDS)	Global and regional climate reanalysis data (ERA5), temperature trends, drought indices, extreme weather	Data sourced from European Centre for Medium-Range Weather Forecasts (ECMWF) and national	Fully open access; free to download after creating an account.	Copernicus Free and Open Data Policy: No restrictions on reuse, including for commercial purposes. Attribution

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.

	indicators. Data format: NetCDF, GRIB.	meteorological agencies. No open submission from individuals.		required.
European Climate Data Explorer (Climate-ADAPT)	Climate indicators for heatwaves, precipitation, and water stress, linked to CDS. Data format: GeoTIFF, CSV, interactive graphs.	Data is curated from EU-funded projects and Copernicus Climate Service; individual scientists and projects can propose datasets through JRC channels.	Fully open access; available for download in multiple formats.	CC-BY 4.0 license; allows free reuse with attribution. Some datasets require acknowledging funding sources.
DRMKC Risk Data Hub	Disaster loss data, climate risk maps (heatwaves, floods, wildfires), vulnerability assessments. Format: GIS layers, PDFs.	Data from EU member states and regional disaster management agencies. No open public upload.	Partial open access; most datasets are open, but some require user registration.	Attribution-based reuse; datasets require citation of original sources. Certain proprietary layers may have limited redistribution rights.
Global Biodiversity Information Facility (GBIF) (Relevant for Ecosystem-based Adaptation)	Observational biodiversity data for climate resilience planning, species migration patterns due to climate change. Format: CSV, Darwin Core.	Open contribution model; any citizen scientist, research group, or institution can upload data following metadata standards.	Fully open access; data are publicly available without restrictions.	CC0 (Public Domain); completely open for reuse, even commercially.
World Meteorological Organization (WMO) Data Platform	Global meteorological and hydrological data, extreme weather alerts, climate	Data contributed by national meteorological agencies, not open for direct public upload.	Controlled access; some datasets are open, others require	Mixed reuse policy; some datasets are free to use, others have restrictions based on

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.

	variability reports. Format: WMO codes, JSON, XML.		institutional access.	country of origin.

The open data ecosystem supporting the Adaptation to Climate Change mission has several notable strengths. Copernicus CDS and Climate-ADAPT serve as highly reliable, fully open-access repositories providing climate projections, historical trends, and extreme weather indicators. Their compliance with FAIR data principles ensures that climate datasets are well-structured, interoperable, and suitable for integration into adaptation models. DRMKC Risk Data Hub further strengthens climate adaptation efforts by offering impact-focused data, such as flood and wildfire risk maps, which help regions design targeted resilience strategies.

Another major advantage is the presence of GBIF, which brings an ecosystem-based approach to adaptation. Biodiversity data is crucial for understanding how climate change affects species distribution, ecosystem health, and natural adaptation processes. The fact that GBIF allows open contributions from citizen scientists makes it particularly valuable in engaging the public and expanding adaptation knowledge beyond institutional datasets.

Most repositories also follow clear and permissive reuse policies. The Copernicus Free and Open Data Policy and CC0 licensing of GBIF ensure that stakeholders—including governments, researchers, NGOs, and businesses—can freely use climate adaptation data without legal or financial barriers. These policies encourage innovative applications, from predictive modelling to climate risk assessments, enabling broader participation in climate resilience planning.

Despite these strengths, several challenges and gaps exist in the current adaptation data infrastructure. One of the key issues is limited direct citizen participation in most repositories. While GBIF is the only major platform that allows open contributions, Copernicus CDS, Climate-ADAPT, and DRMKC primarily rely on institutional data sources. This restriction can exclude valuable local knowledge, especially from community-driven adaptation efforts such as grassroots climate monitoring and indigenous climate observations.

Another limitation is data fragmentation and accessibility constraints. While some repositories offer highly structured datasets, others, such as WMO Data Platform, impose access restrictions on certain meteorological data. This makes it harder for local organisations and researchers without institutional affiliations to fully leverage global climate datasets. Even when datasets are technically open, variations in metadata formats and classification systems can hinder interoperability across different repositories, reducing the efficiency of cross-sector climate modelling.

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.

Moreover, disaster risk data is still not fully open or harmonised. DRMKC Risk Data Hub provides valuable climate risk maps, but some datasets require user registration, and licensing terms vary depending on the original data provider. This creates a challenge for integrating climate adaptation planning with disaster response strategies, as users must navigate multiple legal and administrative hurdles before gaining full access.

Finally, long-term sustainability remains a concern. Several repositories, particularly those linked to specific EU-funded projects, may face funding uncertainties once their initial grants expire. Without clear commitments for continuous maintenance, datasets may become outdated, limiting their usefulness for long-term adaptation planning. Ensuring repository longevity is crucial to keeping climate adaptation data accurate and accessible over time.

To conclude, the open data landscape for the Adaptation to Climate Change mission is strong in providing authoritative, high-quality datasets, particularly through Copernicus CDS and Climate-ADAPT, but it falls short in citizen participation and accessibility consistency. While reuse policies are generally open and encouraging, barriers to data submission, access restrictions on certain platforms, and fragmentation between climate and disaster risk datasets weaken the overall effectiveness of these repositories. Addressing these challenges will be essential for ensuring that open data truly serves as a pillar of climate resilience and adaptation efforts across Europe.

4.2. Cancer

The EU Mission 'Cancer' aims to improve the lives of over 3 million people by 2030 through prevention, early detection, treatment, and better quality of life. Supporting Europe's Beating Cancer Plan, it promotes innovative solutions and equitable care across Member States. Achieving this requires accessible, standardised, and interoperable health data. Open data repositories enable cross-border research, support new diagnostic tools, and facilitate collaboration. Citizen science is also gaining relevance, particularly in patient-reported outcomes and health data from mobile and wearable devices.

Importance of open data repositories for cancer research and prevention

Open repositories enhance transparency, reproducibility, and innovation in cancer research. They allow for the large-scale analysis of imaging, genomic, and clinical data, while also facilitating collaborative development of diagnostic tools—especially those leveraging AI and machine learning. Notably, repositories like *PatientsLikeMe* enable individuals to report and track health outcomes, providing valuable longitudinal data that is difficult to capture in clinical settings.

However, the degree of openness varies significantly across platforms, often due to the sensitive nature of medical data. Ethical safeguards, such as anonymisation and usage

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.

licenses, are essential. At the same time, they may limit broader accessibility or public participation. Striking a balance between privacy and openness is therefore central to cancer-related data governance. Table 5 maps out digital repositories that deal with cancer-based data on the 4 criteria.

Table 5: Digital data repositories on cancer

Repository	Types of Data Shared	Requirements for Data Submission	Policy for Data Access	Policy for Data Reuse
The Cancer Imaging Archive	Radiological images (MRI, CT, PET), metadata	Open to researchers; application and relevance review required	Free access; registration required; anonymisation mandatory	Mostly under CC BY 3.0 / 4.0; some collections may restrict commercial use
PatientsLikeMe	Self-reported patient outcomes, symptom tracking	Only individuals can submit; no institutional bulk upload	Partially open; access to aggregated data limited	Not fully transparent; data reuse policy not clearly stated
Genomic Data Commons (GDC)	Genomic sequences, gene expression, clinical data	Institutional affiliation and Data Use Agreements (DUAs) required	Tiered access: some open, others require application and approval	Varies by dataset; some open-use, others bound by DUA and project-based conditions
European Genome-phenome Archive (EGA)	DNA, RNA sequencing, phenotype information	Institutional submission via Data Access Committees (DACs)	Controlled access; requires ethics approval for use	Restricted reuse; each new use typically requires re-application and consent

Cancer-specific repositories are generally strong in data quality, curation, and adherence to FAIR principles. Platforms like TCIA offer accessible, well-structured imaging datasets, while GDC and EGA serve as foundational infrastructures for genomics. However, despite their robustness, these repositories reveal several challenges. Strict access requirements and institutional submission policies limit participation from citizen scientists and smaller research groups. Patient-led platforms such as PatientsLikeMe

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.

provide valuable decentralised data but often lack transparency in access protocols and interoperability with larger data ecosystems.

Moreover, while many repositories support open access to some extent, licensing conditions vary significantly, often limiting commercial use or secondary applications. The absence of harmonised metadata standards and inconsistent reuse policies can further hinder collaborative research and slow integration across datasets. These barriers reduce the potential for citizen contributions and public engagement—key elements in fostering an inclusive approach to cancer research under the EU Mission framework.

Strengthening citizen science participation will require the development of privacy-preserving models for data sharing, greater clarity on reuse policies, and a shift toward interoperable platforms that support ethical yet open access. By addressing these structural gaps, cancer data repositories can become more inclusive and responsive to the mission's dual goals of scientific excellence and societal impact.

4.3. Restore our Ocean and Waters by 2030

The EU Mission to 'Restore our Ocean and Waters' aims to *'protect and restore the health of our ocean and waters through research and innovation, citizen engagement and blue investments'*. The mission has three specific objectives:

i. *Protect and restore marine and freshwater ecosystems and biodiversity, in line with the EU Biodiversity Strategy 2030 II.*
ii. *Prevent and eliminate pollution of our ocean, seas and waters, in line with the EU Action Plan towards Zero Pollution for Air, Water and Soil III.*
iii. *Make the sustainable blue economy carbon-neutral and circular, in line with the proposed European Climate Law and the holistic vision enshrined in the Sustainable Blue Economy Strategy*

Open data is foundational to achieving the three objectives of the EU Mission to 'Restore our Ocean and Waters'. By providing reliable, interoperable, and openly accessible datasets on biodiversity, pollution, and economic activity, open data enables monitoring of marine and freshwater ecosystems (Objective I), identification of pollution sources (Objective II), and innovation in sustainable, low-carbon blue economy sectors (Objective III). While open data supports all three goals, it is especially critical for objective I, as effective ecosystem protection and restoration rely heavily on timely, granular, and spatially explicit data to guide interventions and measure ecological change. Repositories in this domain consolidate vital information ranging from seabed mapping and oceanographic indicators to human activity records and species tracking data. These underpin marine spatial planning, water quality assessments, and biodiversity restoration strategies. Moreover, citizen science initiatives such as water monitoring and species sightings contribute decentralised, real-time data that complements institutional efforts, fosters transparency, and empowers researchers, policymakers, and

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 public to make informed decisions and track progress collectively. Table 6 maps out digital repositories that deal with marine data based on the 4 criteria.

Table 6: Digital data repositories on freshwater and ocean water

Repository name	Type of data shared	Requirement for data submission	Policy for data access	Policy for data reuse
EMODnet	Marine data across seven domains: bathymetry, geology, physics, chemistry, biology, habitats, human activity	Submission involves structured metadata and documentation; coordinated via assigned national data centres	Free to access and download	Free to reuse with attribution; supports open science
Marine Data Archive (MDA)	General marine datasets, unrestricted by discipline	Managed by data provider; not all content is publicly listed; access based on folder-level permissions	Partially open; public folders accessible, others restricted	Reuse subject to permission from intellectual property holders
SeaDataNet	Oceanographic and marine environmental data	Contributors must be registered institutions or recognised marine data centres	Controlled access; some datasets require authentication	Varies by dataset; often requires licence agreements and attribution
iNaturalist (Marine)	Citizen-contributed species sightings (marine fauna and flora)	Open to individuals via mobile/web platforms; data verified by peer community	Fully open access; data downloadable in bulk or by API	Creative Commons licences (CC-BY or CC0); supports unrestricted reuse with attribution
Copernicus Marine Service (CMEMS	Oceanographic data from satellites and in situ observations	Managed by Copernicus data providers	Free and open access; registration may be required	Re-use permitted under Copernicus license with attribution
Ocean	Marine species	Submission by	Fully open	Creative

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.

Biodiversity Information System (OBIS)	distribution data	registered users with quality control	access	Commons (CC-BY); requires attribution
PANGAEA	Earth and environmental sciences, including marine data	Peer-reviewed datasets with detailed metadata	Open access with DOI assignment	Permitted under open licenses; attribution required
FreshWater Watch	Freshwater quality data from citizen science	Open to individuals via mobile/web platforms; validated by Earthwatch	Data access upon registration; partly open	Reuse permitted for non-commercial use with attribution
GLEON	Lake ecosystem and sensor data	Submission by partner observatories and researchers	Open access for most datasets	Requires attribution; may vary by dataset
GRDC (Global Runoff Data Centre)	Global river discharge data	Submitted by national hydrological services	Free access with registration	Subject to GRDC terms; attribution required
HydroShare	Hydrological data and models	Open submission by researchers and institutions	Open access; supports collaborative sharing	Creative Commons licences; attribution required
AQUASTAT (FAO)	Global water use and agricultural irrigation data	Managed by FAO with national data inputs	Fully open access	Free reuse with attribution

Marine and freshwater data repositories such as EMODnet, CMEMS, and SeaDataNet are central to ocean governance, offering high-quality, interoperable datasets that support research, policy, and the sustainable blue economy. EMODnet stands out for its broad domain coverage and alignment with INSPIRE and FAIR principles, while CMEMS provides real-time oceanographic data critical for forecasting and climate services. SeaDataNet, though rich in institutional datasets, imposes access restrictions that can limit reuse by independent researchers and civil society actors.

Platforms like OBIS and iNaturalist (Marine) demonstrate the growing value of open, community-contributed biodiversity data. OBIS aggregates global species records under open licences, while iNaturalist enables public contributions validated by peers,

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.

particularly useful for tracking biodiversity trends and invasive species. These models enhance inclusivity, but their integration into formal data infrastructures remains partial.

In freshwater domains, repositories such as FreshWater Watch, GRDC, and HydroShare support citizen science and hydrological modelling, yet differences in metadata standards and licensing still hamper interoperability. Controlled access in repositories like GRDC or AQUASTAT can restrict broader use, despite their value.

To conclude, fragmented access, inconsistent reuse conditions, and under-recognition of citizen science limit the potential of open data ecosystems. To advance the EU Mission's goals, data infrastructures must streamline submission, harmonise licensing, and elevate participatory contributions. Strengthening metadata interoperability and supporting hybrid data models will be key to building a more inclusive and responsive knowledge system for ocean and freshwater restoration.

4.4. 100 Climate-neutral and smart cities by 2030

The EU Mission on 'Climate-Neutral and Smart Cities' aims to support 100 European cities in achieving climate neutrality by 2030, serving as innovation hubs and role models for urban sustainability across the continent. This ambitious target encompasses emissions reduction, green mobility, digital transformation, and inclusive governance, calling for a systems-level shift in how cities manage energy, mobility, housing, and public services.

To plan, monitor, and evaluate such transformations, open data plays a pivotal role. Urban data repositories support cities in measuring emissions, optimising energy use, tracking mobility flows, and identifying patterns of environmental injustice. Moreover, citizen science contributes to these efforts by enabling residents to collect hyperlocal data on air quality, noise pollution, energy usage, and transportation patterns, providing granular insights often missed by top-down systems. Table 7 maps out digital repositories that deal with climate neutrality data based on the 4 criteria's.

Table 7: Digital data repositories on climate neutrality

Repository name	Type of data shared	Requirement for data submission	Policy for data access	Policy for data reuse
Smart Cities Marketplace	Project-level datasets on energy, mobility, and sustainability	Submissions limited to EU-funded projects; follows SCIS data guidelines	Open access; downloadable reports and indicators	Reuse terms not clearly specified; varies by dataset and project
Smart Citizen Platform	Sensor-based data on air	Open to individuals; API	Fully open; real-time data	CC-BY or CC0; data can be

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, noise, temperature, etc.	access; integrates with Zenodo for data storage	and historical datasets	reused with attribution
Zenodo (Urban Projects)	Research data, urban trials, mobility logs, etc.	Open participation; widely used by Horizon-funded projects	Open access; multiple formats supported	Creative Commons licences; unrestricted reuse where CC0 or CC-BY is used
OpenStreetMap (OSM)	Crowdsourced geospatial data including roads, infrastructure, public spaces	Open to all; community-led validation and editing	Fully open; data downloadable or accessed via API	ODbL licence; requires attribution and share-alike provisions
KM4City Ecosystem	Urban data from various European cities; supports smart city applications	Open-source platform; data integration from multiple city sources	Open access; data used in dashboards and applications	Data reuse policies depend on individual city contributions
Panoramax	Crowdsourced street-level imagery	Open to individual contributors; images undergo anonymization	Open access to images; API available for integration	Images published under open licenses such as CC BY-SA 4.0

The open data ecosystem supporting the Climate-Neutral and Smart Cities mission reflects a growing recognition of data as a cornerstone for sustainable urban transformation. Platforms like the Smart Cities Marketplace offer structured, project-level data across key sustainability domains including energy, mobility, and digital infrastructure. These repositories help cities showcase progress, benchmark innovation, and facilitate knowledge transfer between urban initiatives across Europe. Zenodo, in turn, supports open data publication for a wide range of urban research outputs, enhancing data visibility and long-term preservation, especially for EU-funded smart city trials and urban labs.

A notable strength of this ecosystem is the integration of bottom-up, sensor-based citizen science contributions, particularly through the Smart Citizen Platform. By enabling individuals to monitor environmental conditions such as air quality and noise pollution, this platform supports real-time, hyperlocal insights into urban sustainability challenges. The seamless integration with Zenodo for archiving ensures that these

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.

datasets are not only participatory but also persistent and citable. Similarly, OpenStreetMap's community-curated geospatial data underpins many open urban planning tools, offering critical spatial layers such as cycling infrastructure, green space distribution, and accessibility maps.

Most repositories in this domain follow permissive reuse policies, encouraging innovation across both civic tech and academic research. Platforms such as Zenodo and Smart Citizen operate under Creative Commons licences (CC0 or CC-BY), allowing datasets to be reused with minimal restrictions. This licensing openness supports data interoperability and encourages novel applications in fields such as digital twins, emissions tracking, and urban equity mapping.

However, significant challenges remain in achieving the full potential of these repositories. One key limitation is the inconsistency in data quality and standardisation, particularly across citizen-generated datasets. While grassroots contributions provide valuable local insights, they often lack harmonised metadata or validation protocols, limiting their integration into formal city planning frameworks. Conversely, institutional repositories such as the Smart Cities Marketplace may prioritise high-level indicators but offer limited granularity or interoperability across cities.

Another concern is the fragmented nature of data access and submission policies. Some platforms, like the Marketplace, restrict data uploads to EU-funded initiatives, excluding contributions from NGOs, citizen groups, or unfunded grassroots experiments. This constraint limits the inclusivity and completeness of the urban data landscape, especially in less resourced cities or regions.

Moreover, the sustainability of participatory platforms remains uncertain, particularly when tied to time-bound projects or reliant on voluntary maintenance. Without long-term institutional support or integration into municipal systems, the longevity of citizen science platforms may be compromised, threatening the continuity of valuable urban datasets.

The open data infrastructure for the Smart Cities mission is advancing rapidly, with promising examples of both institutional and citizen-driven contributions. Yet, to fully realise the mission's vision, greater attention is needed to improve data interoperability, standardise quality protocols, and integrate citizen science into mainstream urban governance. Doing so would ensure that smart city development remains not only data-rich and technologically advanced but also inclusive, democratic, and responsive to the lived realities of urban residents.

4.5. A Soil Deal for Europe

The EU Mission 'A Soil Deal for Europe' aims to establish 100 living labs and lighthouses to lead the transition towards healthy soils by 2030. Soils are foundational to food

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.

security, climate mitigation, water regulation, and biodiversity conservation, yet over 60% of Europe's soils are currently considered unhealthy (European Commission, 2024). The Mission seeks to reverse this degradation by promoting sustainable land use practices, monitoring soil health indicators, and engaging citizens and stakeholders in co-developing and testing solutions through place-based experimentation. To achieve this, the mission set out 8 specific goals:

i. Combat desertification by halting land degradation and achieving restoration, aligned with SDG 15.3.
ii. Increase soil organic carbon stocks by reversing losses on cultivated land and restoring peatlands and wetlands to act as carbon sinks.
iii. Eliminate net soil sealing by increasing land recycling and reducing new urban land sealing to meet "no net land take" goals by 2050.
iv. Reduce soil pollution and enhance restoration by halving pesticide and fertiliser use, promoting organic farming, reducing salinisation, and improving soil health overall.
v. Prevent erosion by reducing the area affected by unsustainable soil water erosion from 25% to sustainable levels.
vi. Improve soil structure and compaction to enhance habitat quality for biodiversity and crop productivity.
vii. Lower the EU's global soil footprint by aligning imports of food, timber, and biomass with sustainability and reducing their impact on global land degradation.
viii. Boost soil literacy and citizen engagement through education, participation, and increased awareness of soil's societal value among EU citizens and stakeholders.

Open data repositories are central to achieving the eight objectives of the EU Mission '100 Living Labs and Lighthouses to Lead the Transition towards Healthy Soils by 2030.' They enable the collection, standardisation, and sharing of diverse soil-related datasets—including soil profiles, nutrient levels, organic carbon content, erosion risks, compaction, and land-use change—across scientific, policy, and community stakeholders. Open data contributes most directly to operational objective 1 and objective 3 by offering spatially explicit and longitudinal data to monitor degradation and urban expansion. More broadly, these repositories support all eight objectives by facilitating evidence-based action on soil carbon restoration, pollution reduction, erosion control, structural improvement, footprint reduction, and public awareness. Citizen science also plays a growing role, particularly through participatory initiatives with farmers, schools, and communities who contribute localised observations and samples. Interoperable, accessible data platforms thus provide the connective infrastructure for aligning scientific research, policymaking, and grassroots action toward the shared goal of healthy soils. Table 8 outlines digital repositories dealing with soil health based on the 4 criteria.

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 8: Digital repositories on soil health

Repository name	Type of data shared	Requirement for data submission	Policy for data access	Policy for data reuse
European Soil Data Centre (ESDAC)	EU-level datasets on soil properties, land use, erosion, organic matter, etc.	Submissions not open to individuals; primarily via EU institutions and projects	Open access to most datasets	Reuse allowed with attribution; licenses vary by dataset (often CC-BY or similar)
ISRIC – World Soil Information	Global soil profiles, maps, and soil property data	Data submission via email or URL; accepts analogue/digital reports	Freely accessible through the WoSIS portal	Varies; users must comply with original data source license (CC-BY or CC-BY-NC)
LUCAS Soil (Eurostat)	Soil survey data across the EU, including chemical and physical properties	Data collected via systematic EU surveys; no public submission option	Downloadable via Eurostat database	Public domain; free reuse with source attribution
OpenLandMap	Global maps on soil pH, texture, organic carbon, erosion, etc.	Open contribution via GitHub for code; data derived from multiple sources	Open access through GeoTIFF & NetCDF formats	CC BY 4.0 license; free reuse with attribution

The open data infrastructure for the 'Soil Deal' mission is supported by a mix of EU-led, global, and community-driven platforms. ESDAC remains the flagship repository for soil data in Europe, offering authoritative datasets that inform everything from Common Agricultural Policy design to erosion risk modelling. Its strength lies in harmonised metadata standards, long-term data stewardship, and robust institutional backing via the Joint Research Centre (JRC). Complementing ESDAC, ISRIC's World Soil Information service curates global soil profile data through its WoSIS platform, integrating datasets from research institutes, national surveys, and scientific publications.

One of the mission's strengths is the increasing granularity and spatial resolution of available soil data, enabled by platforms like OpenLandMap. These repositories provide

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.

open geospatial layers that are machine-readable, compatible with GIS tools, and regularly updated. This enhances accessibility for researchers, policymakers, and practitioners. Moreover, OpenLandMap's use of open-source modelling and transparency in methodology aligns well with FAIR principles, allowing others to reproduce and refine their maps.

Citizen engagement in soil data collection is still emerging but shows strong potential. Initiatives such as school-led soil sampling campaigns and farm-based monitoring trials are increasingly recognised within the living labs framework. However, major repositories like ESDAC and LUCAS do not yet enable direct data submission from citizens or civil society actors, creating disconnect between grassroots efforts and official data infrastructures.

Another challenge lies in the reuse conditions across platforms. While most datasets are freely accessible, licenses and attribution requirements vary. For example, ISRIC's decentralised model means that datasets can carry different terms, including non-commercial restrictions. This inconsistency may complicate integration efforts or limit reuse in applied settings such as agritech development or regenerative agriculture platforms.

Furthermore, some repositories, especially LUCAS and ESDAC, lack APIs or user-friendly tools for real-time data interaction, which can restrict their usability for citizen-led monitoring or dynamic land-use modelling. Improving technical interoperability—such as enabling RESTful APIs, real-time data submission, or modular dashboard development, would help bridge this usability gap.

Finally, long-term repository sustainability and cross-institutional coordination remain essential. Soil data repositories are often maintained within large institutional silos, with limited formal pathways for integrating academic, private sector, and community datasets. Developing shared standards for soil health indicators and open-source frameworks for federated data sharing will be vital to creating a resilient, future-ready soil data ecosystem.

Overall, while the Soil Deal mission benefits from a mature and scientifically rigorous data landscape, there is significant room to expand inclusivity, interoperability, and civic participation. Enhancing the openness of submission channels, clarifying licensing terms, and investing in participatory infrastructure would empower citizens and practitioners to co-lead Europe's transition toward healthy soils.

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.

5. CROPS and ScienceUs project practices

ScienceUs[3] is a Horizon Europe-funded "sister" initiative to the CROPS project, working to promote and institutionalise citizen science across Europe. The project focuses on strengthening localised citizen engagement in scientific research and ensuring that community-generated knowledge is recognised within policymaking and climate adaptation strategies. Like CROPS, ScienceUs is grounded in the principles of open science and transdisciplinary collaboration.

What makes ScienceUs particularly relevant and complementary to CROPS is its emphasis on capacity-building for grassroots science initiatives. Through mentoring, transnational networking, and structured support, ScienceUs helps small-scale citizen science projects scale their impact, communicate their findings more effectively, and share data in transparent, accessible formats.

5.1. The 25 Projects Identified: Purpose, Progress, and Positioning

Under its SEED Phase, ScienceUs funded 25 pilot projects across Europe that span a wide spectrum of thematic areas including environmental monitoring, sustainable mobility, biodiversity protection, genomic research, and participatory education. These projects were selected for their innovative approaches to citizen engagement and their potential to contribute meaningful data to both local and European climate and sustainability goals.

The primary aim of these projects is to test, adapt, and scale citizen science methodologies within real-world contexts. Each project is required to produce a *transnational upscaling and campaign plan*, ensuring that their methods and findings can be reused or replicated beyond their immediate locality. In doing so, ScienceUs lays the groundwork for long-term sustainability and integration of citizen science into mainstream policy frameworks.

This set of projects provides a unique testbed for understanding how citizen science efforts approach data sharing—a key pillar of both ScienceUs and CROPS' missions. These initiatives not only gather locally relevant data but are also expected to disseminate it in ways that align with open science principles.

5.1.1. Type of data shared

The ScienceUs-funded projects showcase a wide variety of data types, reflecting the interdisciplinary nature of the initiative:

- **Genomic and Biomedical Data**: Shared by research-intensive institutions such as the Centre for Genomic Regulation (CRG) and the Melanogaster project. These

[3] https://scienceus-project.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.

datasets are typically subject to strict access controls due to ethical and privacy considerations.

- **Environmental and Marine Data**: Provided by institutes like CIIMAR, this data contributes to understanding biodiversity, marine ecosystems, and supports broader environmental conservation efforts.
- **Citizen-Generated Data**: Exemplified by the SCORE Smart Pebbles Workshop, this type of data is collected through participatory science methods and focuses on local issues such as coastal erosion monitoring.
- **Academic Research Outputs**: Includes publications, teaching materials, and datasets, typically hosted on institutional repositories like DSpace or Digital Commons. These repositories ensure both preservation and accessibility of scholarly outputs.

5.1.2. Openness and accessibility

The degree of openness in data sharing varies significantly across projects:

- **Open Access**: Several projects use open-access repositories like Zenodo or university-managed platforms, providing unrestricted public access. This supports transparency, collaboration, and reusability.

- **Controlled Access**: Projects dealing with sensitive data, particularly in biomedical fields, rely on secure repositories such as the European Genome-Phenome Archive (EGA). Access is granted under ethical review or formal data-use agreements.

- **Unspecified or Limited Access**: A number of project websites do not clearly describe their data-sharing mechanisms or repository use. This limits visibility, hinders interoperability, and poses a challenge for replicability and cross-project learning.

Table 9 provides a summary of the above text by mapping out the ScienceUs funded project based on their data type and accessibility.

Table 9: Mapping of ScienceUs projects based on data type and accessibility

Type of access	Name of ScienceUs funded Project	Data type	URL
Open Access	Centre for Genomic Regulation (CRG)	Genomic and Biomedical Data	https://www.crg.eu/
	Melanogaste	Genomic and	https://melanogaster.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.

	r: Catch the Fly!	Biomedical Data	
	University of Patras	Academic Research Outputs	https://www.upatras.gr/en/
	Universitatea Babeş-Bolyai	Academic Research Outputs	https://www.ubbcluj.ro
	National Institute of Biology (NIB)	Genomic and Biomedical Data	https://www.nib.si/eng/
	SCORE Smart Pebbles Workshop	Citizen-Generated Data	https://score-eu-project.eu/
Controlled Access	Centre for Genomic Regulation (CRG)	Genomic and Biomedical Data	https://www.crg.eu/
Unspecified/Limited access	SciencePost	Academic Research Outputs	https://www.sciencepost.pt/
	Tatavaka	Citizen-Generated Data	https://tatavaka.hr/
	CIIMAR	Genomic and Biomedical Data	https://www.ciimar.up.pt/
	LC3	Environmental and Marine Data	https://lc3-nzlimassol2030.eu/en/
	DemosLab	Academic Research Outputs	https://www.demoslab.com/
	Germinando	Academic Research Outputs	https://germinando.es/
	Blue Baltic Ecosystem	Environmental and Marine Data	https://bluebalticecosystem.com/
	EDU-MOVE – Tirana	Citizen-Generated Data	https://accting.eu/selected-pilot-projects/

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.

	Università Roma Tre	Academic Research Outputs	https://www.uniroma3.it/
	Placemaking Europe	Academic Research Outputs	https://www.placemakingweb.org
	Teva	Environmental and Marine Data	https://www.teva.org.il/
	Science for Inclusion	Academic Research Outputs	http://scienceforinclusion.it/
	Associació Taxus	Environmental and Marine Data	https://associaciotaxus.cat/lassociacio/
	Asociación Hippocampus	Genomic and Biomedical Data	https://asociacionhippocampus.com
	IoT4Nature	Citizen-Generated Data	https://www.iot4nature.ro
	Fundación Vida Sostenible		https://www.vidasostenible.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.

6. Social media for citizen science data dissemination

Beyond the structured support offered by programmes like ScienceUs, a diverse array of citizen science initiatives across Europe and globally have developed their own methods of sharing data and engaging the public—often through informal channels such as social media, blogs, and interactive community platforms. These projects may not be ScienceUs-funded, but they reflect the same commitment to openness, participation, and grassroots knowledge generation. By leveraging platforms like Instagram, Facebook, and mobile apps, these initiatives make scientific data more visible, relatable, and actionable—particularly for communities not typically reached by institutional science. Exploring how these projects communicate findings and share citizen-generated data offers useful insights for the broader citizen science ecosystem, including CROPS and similar initiatives aiming to foster decentralised, publicly meaningful science. The following examples illustrate how citizen science projects are using these channels to share data, engage participants, and build broader public awareness

Table 10: Repositories supporting citizen-generated local data collection

Project/Repository Name	Platform	What is shared	How it's shared
iNaturalist (Global / EU Users)	Mobile app, website, Twitter, Instagram	Geo-tagged biodiversity observations submitted by the public (e.g., birds, insects, plants)	Public display of photos and location-based observations via the app and website Aggregation into global biodiversity databases such as GBIF Social media used to feature "species of the week," tutorials, and project milestones
FreshWater Watch (Earthwatch Europe)	Website, social media	Water quality, turbidity, and pollution data collected by citizen	Summarised results presented through interactive dashboards and maps Awareness

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.

		volunteers	campaigns timed with events such as World Water Day Posts featuring citizen-collected data, short videos, and field photos
EU-Citizen.Science (EU Horizon 2020)	Web portal, newsletter, Twitter, LinkedIn	Project metadata, tools, results, and methodologies (but not raw data)	Highlights of best practices and toolkits across social channels Cross-posting of updates and case studies from participating projects Promotion of training resources and community events.
Observadores del Mar (Spain)	Website, Instagram, Facebook, YouTube	Marine sightings, coastal changes, and plastic pollution data	Volunteer-submitted posts documenting marine life or pollution Data verified by scientists and incorporated into marine monitoring databases Awareness-building through infographics, storytelling, and community competitions
OPAL – Open Air Laboratories (UK)	Website, Twitter, Facebook (project now archived; legacy continues)	Soil health, air quality, tree health, and biodiversity data from schools and individuals	Community findings shared via blog posts, interactive maps, and downloadable reports Seasonal citizen campaigns (e.g., tree surveys) used to drive

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.

			participation Teachers and school groups featured prominently in dissemination materials

Common Strategies Observed Across Projects

Across diverse citizen science initiatives, several shared strategies have emerged for communicating findings and mobilising participation through informal digital channels. These practices illustrate how projects translate citizen contributions into widely accessible, engaging, and policy-relevant outputs:

- Data Visualisation and Storytelling: Projects often transform raw data into accessible formats, such as infographics, interactive dashboards, or story maps, to enhance public understanding and encourage participation. FreshWater Watch, for instance, presents citizen-gathered water quality data through interactive maps and visual campaigns aligned with events like World Water Day, making complex findings more relatable.

- Participant recognition and community building: Featuring individual contributors fosters a sense of ownership and encourages sustained engagement. iNaturalist highlights community members' observations through regular "species of the week" posts and user-credited images on platforms like Instagram and Twitter, reinforcing collective identity and visibility.

- Cross-platform dissemination: Content is frequently adapted across various platforms to maximise reach and resonance. Observadores del Mar, for example, shares marine sightings and plastic pollution reports through Instagram, Facebook, YouTube, and its website, tailoring the message and format to suit each medium.

- Integration with formal infrastructures: Many projects connect informal citizen-collected data to formal scientific or policy databases. Biodiversity records submitted via iNaturalist are routinely integrated into the Global Biodiversity Information Facility (GBIF), bridging grassroots observation with institutional research.

- Thematic campaigns and event-driven outreach: Strategic alignment with environmental observances boosts visibility and public participation. FreshWater Watch times its outreach efforts with international awareness days, combining citizen stories, short videos, and data highlights to expand engagement.

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.

These approaches offer valuable insights for CROPS and similar initiatives aiming to upscale decentralised citizen science, particularly by enhancing visibility, fostering public ownership, and ensuring that citizen-generated data contributes meaningfully to policy-relevant knowledge systems.

7. Comparative Analysis

This section presents the core comparative analysis of open data repositories relevant to the five EU Missions. The aim is to assess how these repositories perform across key dimensions of openness, accessibility, and citizen science readiness. The analysis identifies patterns, strengths, and gaps in current platforms, with the goal of informing efforts to make open data systems more inclusive, interoperable, and mission aligned.

The section is structured around five analytical dimensions: FAIR compliance, participation pathways, governance, usability, and alignment with EU policy frameworks; each of which reflects criteria central to supporting decentralised, citizen-generated data ecosystems. These dimensions serve as the basis for evaluating selected repositories and drawing cross-cutting insights applicable to initiatives like CROPS.

7.1. Aim

This section conducts a comparative analysis of the open data ecosystems supporting five EU Missions: Adaptation to Climate Change, Cancer, Restore our Oceans and Waters, Climate-Neutral and Smart Cities, and A Soil Deal for Europe. The comparison is structured around five core dimensions that are fundamental to open science, citizen engagement, and sustainable data governance. These dimensions not only reflect alignment with the FAIR data principles (Findable, Accessible, Interoperable, and Reusable) but also indicate the extent to which each mission fosters inclusive, transparent, and participatory science-policy interfaces. The five dimensions are defined as follows:

- **Repository Openness***:* The degree to which mission-related data repositories are open for public access, browsing, and download without restrictive authentication barriers.
  Assessment indicators: Public visibility, ease of access, presence of open APIs, and availability of metadata.

- **Citizen Data Submission**: The extent to which citizens and non-institutional actors can contribute data to mission repositories, including through citizen science, participatory monitoring, or community mapping.
  Assessment indicators: Submission portals, documentation for public contribution, use of citizen science platforms (e.g. EU-Citizen.Science), and policies on data validation.

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.

- **Interoperability**: The capability of repositories to integrate with other platforms and datasets through standardised data formats, vocabularies, and protocols. *Assessment indicators:* Use of metadata standards (e.g. INSPIRE, DCAT), compatibility with EU open data infrastructure (e.g. European Open Science Cloud), and ability to support machine-readable formats.

- **Reuse Licensing Clarity:** The transparency and standardisation of data licensing that governs the reuse, redistribution, and modification of data. *Assessment indicators:* Use of recognised open licenses (e.g. Creative Commons), clear terms of use, and guidance on attribution or reuse conditions.

- **Institutional Sustainability:** The long-term viability of the data repositories, including financial continuity, institutional backing, and mechanisms for periodic updates and community engagement.
  Assessment indicators: Funding models, involvement of national or EU-level institutions, regularity of data updates, and presence of stakeholder governance structures.

These dimensions capture how effectively each mission aligns with FAIR data principles (Findable, Accessible, Interoperable, and Reusable) and how well they support inclusive and sustainable data governance. Together, these dimensions enable a critical understanding of how the Missions operationalise open data strategies and engage diverse actors in mission-oriented research and innovation. The following sub-sections will assess each EU Mission against these dimensions to identify good practices, bottlenecks, and areas for improvement.

7.2. Method

Each mission was scored on a scale of 1 to 3 for each core dimension:

Score	Description
3 (High)	Fully aligned with FAIR principles and open science practices
2 (Moderate)	Some progress, but limited by structural or legal barriers
1 (Low)	Minimal openness, restricted reuse, or exclusionary practices

Scoring Methodology:
Scores were assigned based on a desk-based review of publicly available evidence, including:

- Repository documentation (e.g., terms of use, submission guidelines, metadata schemas)

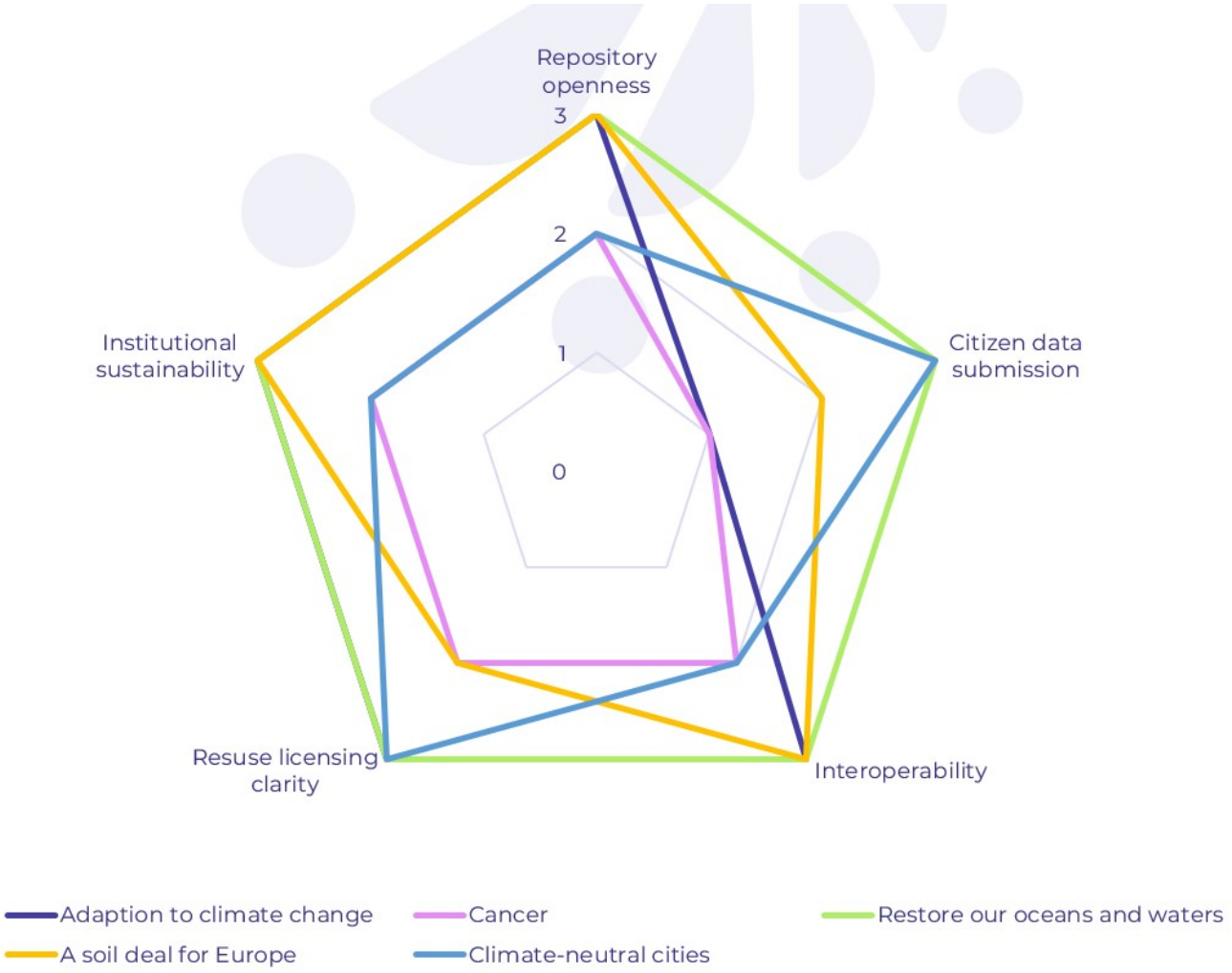
Funded by the European Union

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.

- Data platforms and portals (e.g., Zenodo, Smart Citizen, Data.europa.eu)
- Licensing frameworks (e.g., Creative Commons, Open Database License)
- Citizen science and participatory tools (e.g., iNaturalist, EU-Citizen.Science)
- EU programme outputs (e.g., Horizon project deliverables, technical briefs)
- Interoperability protocols and metadata standards (e.g., INSPIRE, DCAT-AP)

This qualitative assessment aimed to capture both the formal infrastructure (e.g., technical design, licensing clarity) and practical implementation (e.g., actual use cases, community engagement) underpinning each Mission's data ecosystem. Cross-referencing across platforms ensured triangulation of sources and consistency in scoring.

7.3. Results



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.

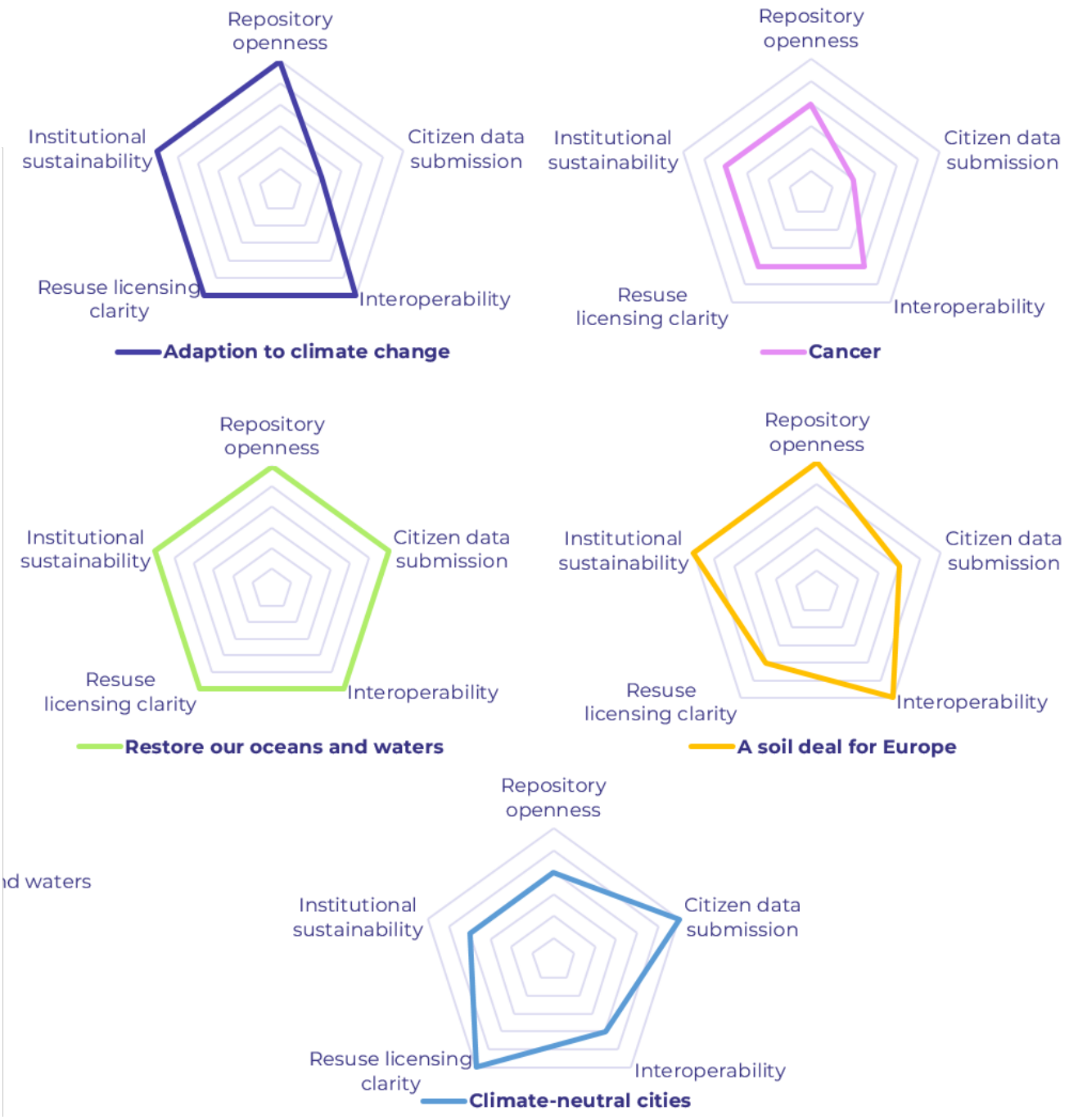


Figure 2: Radial charts of mission scores across five core dimensions (comparative and per each mission)

Figure 2 shows clear contrasts between missions 'Restore our Oceans and Waters' demonstrates consistently strong performance across all five criteria, particularly due to its use of EMODnet and participatory tools like iNaturalist. 'Adaptation to Climate Change' performs well in openness, interoperability, and sustainability, but shows minimal citizen contribution. The 'Cancer' mission scores lowest across most categories, largely due to the ethical and legal constraints inherent in health data governance.

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.

'Smart Cities' scores highly on citizen data submission but suffers from fragmentation in interoperability and unclear reuse licensing. 'A Soil Deal for Europe' shows robust openness and sustainability but lacks citizen contribution mechanisms and clear reuse protocols.

Table 11: Repository-level comparison across the five analytical dimensions

Mission	Repository name	Accepts citizen-submitted data	Is open for use	Is institutionally sustainable	Is interoperable	Has clear reuse licensing
Adaption to climate change	Copernicus CDS	N	Y	Y	Y	Y
	Climate-ADAPT	N	Y	Y	Y	Y
	DRMKC Risk Data Hub	N	Y	Y	Y	N
	GBIF	Y	Y	Y	Y	Y
	WMO Data Platform	N	Y	Y	N	N
Cancer	The Cancer Imaging Archive	N	Y	Y	N	Y
	PatientsLikeMe	Y	N	N	N	N
	Genomic Data Commons (GDC)	N	Y	Y	Y	N
	European Genome-phenome Archive (EGA)	N	Y	Y	N	N
Restore our Ocean and Waters	EMODnet	N	Y	Y	Y	Y
	Marine Data Archive (MDA)	N	N	Y	N	N
	SeaDataNet	N	N	Y	Y	N
	iNaturalist	Y	Y	Y	Y	Y

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.

	Copernicus Marine Service (CMEMS)	N	Y	Y	Y	Y
	OBIS	N	Y	Y	Y	Y
	PANGAEA	N	Y	Y	Y	Y
	FreshWater Watch	Y	Y	Y	Y	N
Climate-neutral and Smart Cities	Smart Cities Marketplace	N	Y	N	N	N
	Smart Citizen Platform	Y	Y	N	Y	Y
	Zenodo (Urban Projects)	Y	Y	Y	Y	Y
	OpenStreetMap	Y	Y	Y	Y	Y
	KM4City	N	Y	N	N	N
	Panoramax	Y	Y	N	N	Y
A Soil Deal for Europe	ESDAC	N	Y	Y	Y	N
	ISRIC – World Soil Information	N	Y	Y	Y	N
	LUCAS Soil (Eurostat)	N	Y	Y	N	Y
	OpenLandMap	Y	Y	Y	Y	Y

Figure 3 presents a repository-level comparison across the five analytical dimensions: citizen data submission, institutional sustainability, interoperability, reuse licensing clarity, and repository openness. The visual highlights significant variation between repositories, revealing both standout platforms and areas of underperformance that would be masked in aggregate mission-level views.

Repositories such as iNaturalist (Marine), GBIF, OpenStreetMap, OpenLandMap, and Zenodo (Urban Projects) emerge as consistently high performing across most or all dimensions. These platforms combine participatory design, technical interoperability, clear reuse licensing, and sustained institutional support, making them strong

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.

exemplars for inclusive and FAIR-aligned open data ecosystems. Notably, several of these repositories also enable direct citizen data submission, reinforcing the value of decentralised, bottom-up contributions in mission-relevant domains.

In contrast, repositories like LUCAS Soil, KM4City, and the Cancer Imaging Archive demonstrate more uneven profiles. LUCAS and KM4City, for instance, score well on institutional sustainability and openness but lack accessible submission pathways and standardised reuse terms. The Cancer Imaging Archive, while robust in ethical governance and structured data, remains constrained by limited participation and strict access controls, a reflection of the broader regulatory barriers associated with health data. Other platforms, such as Smart Citizen, FreshWater Watch, and Panoramax, stand out for their strong citizen engagement but face challenges in areas like interoperability and long-term sustainability, particularly where institutional backing is weak or project based.

This detailed repository-level analysis offers valuable insight for initiatives like CROPS, which aim to upscale decentralised citizen science across Europe. By surfacing the specific features that enable repositories to support openness, participation, and policy relevance simultaneously, the visual provides a roadmap for future alignment. In particular, CROPS can draw on high-performing examples to guide the development of scalable, citizen-inclusive infrastructure, ensuring that grassroots data contributions are not only welcomed but effectively preserved, integrated, and reused within mission-driven research and governance systems.

Discussion of Core Dimensions

The comparative analysis reveals varying degrees of alignment across the five core dimensions, with differences primarily shaped by the maturity, structure, and openness of existing data ecosystems rather than direct outcomes of the Missions themselves. Importantly, open access data is available to support all Missions, even where dedicated Mission-specific repositories or platforms are still under development (e.g., the Soil Mission Dashboard).

- **Repository Openness:** All five Missions benefit from existing repositories that provide open access to at least part of their data. In particular, *Adaptation*, *Oceans*, and *Soil Deal* benefit from mature infrastructures like Copernicus, EMODnet, and ESDAC, which offer openly downloadable datasets under standardised terms. For *Cancer* and *Smart Cities*, data access tends to be more fragmented, with many repositories hosted on project-specific or institutional platforms that may limit public access or require registration.

- **Citizen Data Submission:** This is the most unevenly developed dimension. Tools such as iNaturalist, Smart Citizen, and OpenStreetMap are actively used in *Oceans* and *Smart Cities*-relevant initiatives to integrate citizen-generated data. In

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.

contrast, data submission opportunities in *Cancer* and *Adaptation*-related contexts are generally institutionally managed, with few clear pathways for grassroots or community-level contributions.

- **Interoperability:** Several repositories aligned with *Adaptation*, *Oceans*, and *Soil Deal* make use of recognised metadata standards such as ISO, INSPIRE, or Darwin Core, facilitating integration and discoverability. Conversely, *Cancer*-related platforms often lack harmonisation, with the *Smart Cities* data portals showing a wide variation in format and integration protocols, reflecting their diverse and localised origins.

- **Reuse Licensing Clarity:** Many datasets relevant to *Oceans*, *Adaptation*, and *Smart Cities* apply clear and open reuse licenses—typically Creative Commons—allowing for redistribution and adaptation. However, in repositories linked to *Soil Deal* and *Cancer*, license information is sometimes incomplete, ambiguous, or subject to institutional restrictions, particularly for datasets hosted on project-based platforms.

- **Institutional Sustainability:** Data infrastructures embedded in long-term EU programmes such as Copernicus, EMODnet, and ESDAC benefit from sustained institutional and financial support, providing strong foundations for ongoing access and updates. By contrast, some data platforms used in *Smart Cities* and *Cancer* contexts appear reliant on limited-term project funding, raising concerns about future maintenance and data continuity.

8. Best practices for open data repositories

Drawing from the comparative assessment of repositories across the five EU Missions, this section outlines a set of best practices that, in the reviewers' judgement, contribute to making open data platforms not only FAIR-compliant but also equitable, technically robust, and ethically sound. These practices were identified based on recurring features observed in repositories that most effectively support inclusive participation, sustainable data use, and mission-relevant research and innovation. While the notion of "effectiveness" remains context-dependent, these practices reflect the reviewers' informed perspective on what enables open data ecosystems to function as accessible, interoperable, and socially valuable infrastructures.

a) Use of recognised metadata standards

High-performing repositories consistently adopt domain-relevant metadata standards such as Darwin Core for biodiversity, ISO 19115 for geospatial information, and INSPIRE for harmonised EU spatial data. These standards facilitate cross-platform data discovery and reuse, and in some cases, also support multilingual metadata fields to improve accessibility across the EU. For example, EMODnet, OBIS, and ESDAC exhibit structured

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.

metadata aligned with international or EU-level norms. This consistency enables seamless integration with other repositories, improves dataset findability, and supports long-term curation.

b) Provision of machine-readable formats and APIs

Repositories that offer datasets in machine-readable formats—such as CSV, JSON, NetCDF, or GeoTIFF—enable automation, data visualisation, and integration with modelling or dashboard tools. Many of the leading repositories also provide public APIs, allowing real-time or batch access for research and application development. Platforms like OpenStreetMap, Smart Citizen, Copernicus CDS, and OpenLandMap stand out for offering rich, structured data alongside APIs that facilitate third-party use and innovation.

c) Clear and inclusive reuse licensing

Effective repositories implement clear, open licensing frameworks that allow users to understand and legally reuse the data. Widely adopted licences such as CC0, CC-BY, or ODbL provide permissive terms that promote data sharing, ensure attribution, and reduce legal uncertainty. Repositories such as Zenodo, GBIF, and OpenStreetMap are exemplary in how they display and enforce these licences. Their clarity supports trust and facilitates cross-sectoral uptake—whether in research, policy, or public engagement.

d) Ethical Governance and Participatory Models

Ethical data governance is central to long-term repository trustworthiness. This includes transparency about who manages the data, how decisions about access and use are made, and how contributors are acknowledged. In domains involving sensitive data—such as health, personal mobility, or indigenous knowledge—robust ethical safeguards and review procedures are particularly critical. Repositories like The Cancer Imaging Archive, EGA, and GDC demonstrate rigorous ethical control over access to sensitive datasets. Meanwhile, platforms such as iNaturalist and OpenStreetMap showcase participatory governance, community moderation, and peer validation practices that empower contributors while ensuring data integrity.

e) Equity and Accessibility in Design

Open data platforms should be accessible to a broad range of users, not just researchers in well-funded institutions. Good practices include user-friendly interfaces, multilingual support, and clear contribution pathways for non-experts. For example, Smart Citizen enables sensor-based community input, and EMODnet Biology offers guidance for both institutional and citizen science data submissions. However, accessibility must be balanced with data quality safeguards. Open submission should not mean unrestricted

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.

uploading; effective platforms include validation steps and contribution protocols that enable diverse participation without compromising data integrity.

f) Alignment with Policy and Monitoring Frameworks

Data repositories that align with EU policy instruments or formal indicator systems increase the relevance and usability of their datasets. When open data feeds into frameworks such as the EU Green Deal, SDG monitoring, or the Marine Strategy Framework Directive, it has a direct impact on regulation, funding, and strategic decision-making. This alignment is evident in platforms like EMODnet, which informs marine policy compliance, and ESDAC, whose datasets support monitoring under the Common Agricultural Policy (CAP). These repositories not only store data—they actively support evidence-based governance.

To conclude, the most effective open data repositories combine technical excellence with ethical integrity and inclusive design. The practices outlined in this section reflect the essential elements that allow repositories to serve as critical infrastructure for citizen science, public policy, and transnational collaboration. As EU Missions continue to scale their ambition, such repositories will remain central to ensuring that citizen-generated data is not only collected, but preserved, trusted, and used. For initiatives like CROPS, which aim to upscale decentralised, citizen-driven science, these best practices offer a roadmap for designing and aligning data platforms that are not only FAIR-compliant but also participatory, policy-relevant, and sustainable at scale.

9. Limitations

While this deliverable provides a comparative and cross-dimensional analysis of open data repositories across five EU Missions, several limitations should be acknowledged:

a) Scope of repository selection

The analysis focuses on a curated set of repositories that are either explicitly linked to the EU Missions or widely recognised within their respective domains. However, this selection may not capture the full spectrum of available repositories, particularly smaller-scale or emerging platforms at national or local levels. As a result, some innovative or context-specific practices may be underrepresented.

b) Reliance on publicly available information

Repository evaluations were based primarily on desk research using publicly accessible interfaces, documentation, and platform features. This means internal governance structures, unpublished reuse metrics, or behind-the-scenes data integration efforts may not be fully reflected in the analysis. Moreover, some repositories may have features or governance protocols not visible without registration or institutional affiliation.

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.

c) Scoring simplification

The use of a 1–3 scoring scale across ten dimensions provides a consistent comparative framework but also necessitates some simplification. For instance, nuanced differences between repositories that both score "2" on machine-readability may still be significant in practice. These scores are intended as indicative rather than exhaustive and are best interpreted alongside the qualitative insights provided in each section.

d) Evolving nature of platforms

Open data repositories are dynamic systems, with frequent updates to functionality, licensing models, data types, and API access. Some observations made at the time of analysis may become outdated quickly as platforms evolve. This is particularly relevant for repositories tied to time-bound EU projects, which may lose functionality or visibility after funding ends.

e) Geographic and language biases

Although the deliverable seeks to be EU-wide, there is an inherent bias toward repositories that are accessible in English and/or widely visible through pan-European data portals. This may exclude locally embedded platforms or community repositories that operate in other languages or serve more specific user bases.

f) Exclusion of raw usability and impact data

The analysis does not include empirical metrics of repository usage (e.g. download volumes, contributor demographics, impact on policy decisions) due to limited data availability. As such, it is possible that some high-functioning but low-visibility repositories were under-valued relative to more prominent platforms with better outreach or institutional connections.

10. Recommendations

Drawing on the analysis, best practices, and identified limitations, the following recommendations aim to support the development of more inclusive, interoperable, and sustainable open data ecosystems across the EU Missions.

- Strengthen cross-mission interoperability: Encourage the use of common metadata standards (e.g. ISO 19115, INSPIRE, Darwin Core) and develop crosswalks to enable data exchange between repositories serving different Missions. This would support horizontal integration across thematic areas such as biodiversity ('Oceans' and 'Soil Deal'), climate risk ('Adaptation'), and urban sustainability ('Smart Cities').

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.

- Mandate open licensing and machine-readable formats: Require all EU-funded repositories to publish datasets under open licences (e.g. CC-BY or CC0) and in accessible, machine-readable formats (e.g. CSV, JSON, NetCDF) with documented APIs. This will promote reuse, automation, and integration across both scientific and civic data ecosystems.

- Expand citizen participation pathways: Develop structured submission channels and support tools that enable citizen scientists, schools, and communities to contribute data directly to repositories. These should be accompanied by quality control protocols, validation mechanisms, and contributor guidance to ensure data integrity and usability. Participation should be inclusive but not at the expense of data quality or interoperability.

- Improve governance transparency and user guidance: Ensure repositories publish clear governance frameworks, ethical guidelines, and user support mechanisms for feedback, dispute resolution, and attribution. Transparent governance fosters trust among contributors and users and is particularly vital for repositories handling sensitive or community-generated data.

- Align repositories with policy instruments and monitoring systems: Design repository outputs with explicit links to EU-level monitoring and policy frameworks—such as the Common Agricultural Policy ('Soil Deal'), Marine Strategy Framework Directive ('Oceans'), EU Green Deal ('Smart Cities' and 'Climate Adaptation'), and Europe's Beating Cancer Plan ('Cancer' Mission). This alignment will ensure that open data directly informs regulatory action, strategic investment, and mission performance tracking.

- Support long-term sustainability beyond project lifecycles: Embed data stewardship plans into project designs from the outset, and explore institutional, federated, or mission-driven hosting models to ensure that repositories remain active and reliable beyond the duration of EU funding. This is essential to maintain trust, prevent data loss, and enable cumulative science.

These recommendations provide a practical roadmap for improving the inclusivity, reliability, and long-term impact of open data repositories aligned with EU Mission objectives. By addressing structural gaps in licensing, metadata standardisation, citizen participation, and governance, they support the transformation of fragmented data infrastructures into interoperable, mission-aligned systems. For CROPS and similar initiatives, implementing these actions is essential to upscaling citizen science, ensuring that community-generated data can be effectively integrated, reused, and trusted within both scientific research and policymaking contexts.

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.

11. Extending the framework: cross-cutting analytical dimensions

While the five core dimensions: repository openness, citizen data submission, interoperability, reuse licensing clarity, and institutional sustainability, offer a useful baseline for evaluating open data ecosystems, they primarily focus on technical functionality and structural accessibility. These dimensions assess how well repositories align with FAIR principles and support foundational aspects of open science and inclusive data governance.

However, to fully understand the real-world effectiveness and societal impact of open data systems, future evaluations should also consider cross-cutting analytical dimensions that shape ethical robustness, practical accessibility, and policy relevance. We propose the following additional lenses as essential for more holistic assessments:

11.1. Metadata standardisation

Why it matters: Standardised metadata is essential for making data discoverable, usable, and interoperable. Repositories that use internationally recognised metadata schemas—such as Darwin Core (biodiversity), ISO 19115 (geospatial), or INSPIRE (EU spatial data)—enable data to be easily understood, shared, and reused across disciplines and borders. Well-structured metadata ensures that datasets include critical descriptive elements (e.g., units of measurement, location, time period, data collection method), reducing ambiguity and increasing trust. Moreover, multilingual metadata enhances accessibility across EU member states, ensuring no linguistic group is excluded from using or contributing to the data.
What happens if ignored: Without consistent metadata, datasets may be technically open but functionally inaccessible—difficult to search, compare, or integrate with other sources. This inhibits scientific collaboration, cross-border research, and pan-European policy planning.

11.2. Machine-Readability and Automation

Why it matters: Machine-readable data (e.g. in CSV, JSON, NetCDF formats) and accessible APIs are foundational for modern data-driven analysis, visualisation, and decision-making. They allow researchers, policymakers, and developers to automatically access and process large volumes of data without manual downloading, cleaning, or reformatting. APIs enable real-time integration into dashboards, climate models, and public platforms—allowing data to remain dynamic and continuously relevant.

What happens if ignored: If datasets are only available as PDFs or embedded in static tables, they become hard to reuse or integrate. This blocks innovation, limits policy responsiveness, and prevents the development of interactive citizen tools and applications.

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.

11.3. Equity and Inclusion

Why it matters: Open data should not only be technically accessible, but socially and geographically inclusive. Repositories must be usable by non-expert users, community groups, schools, and organisations from under-resourced regions, not just scientists in well-funded institutions. This includes intuitive interfaces, local language support, and the ability to contribute or engage without technical barriers. It also involves recognising and crediting contributions from citizen scientists, ensuring their role is valued and visible.

What happens if ignored: Repositories that exclude non-expert users or smaller communities reinforce existing inequalities in who can access and benefit from scientific knowledge. Without inclusive design, the promise of citizen science and public participation is undermined. Policy Integration

11.4. Policy Integration

Why it matters: Open data becomes most impactful when it is used in real policy contexts—e.g., in environmental monitoring dashboards, sustainable development indicators, or national adaptation plans. Data that aligns with EU frameworks such as the Green Deal, SDGs, or Common Agricultural Policy becomes more than just a dataset—it becomes actionable evidence that can shape investment, regulation, and public programmes.
What happens if ignored: Without structured pathways for integration into policy systems, citizen-generated or project-based data risks becoming siloed and underutilised. Policymakers may default to traditional data sources, leaving innovative grassroots or interdisciplinary data invisible.

11.5. Governance and Ethical Transparency

Why it matters: Open data is only trustworthy if users understand who controls the repository, how licensing is applied, and how data protection, consent, and attribution are handled. This is especially critical for sensitive domains like health, environmental justice, or indigenous knowledge. Transparent governance means clear terms of use, ethical review processes, and mechanisms for addressing concerns or correcting data. It supports responsible innovation, avoids misuse, and builds long-term confidence among contributors and users.

What happens if ignored: Lack of governance clarity can result in data misuse, legal ambiguity, or reputational harm. It may also discourage citizen contributions due to uncertainty about how their data will be handled or whether their privacy will be respected.

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.

12. Conclusion

This research has shown that while many open data repositories across the five EU Missions exhibit strong technical infrastructure and alignment with FAIR principles, their capacity to support inclusive, citizen-driven science remains uneven. Repositories such as EMODnet, Copernicus, and ESDAC stand out for their interoperability, metadata quality, and policy relevance. However, most platforms still lack clear pathways for citizen data submission, consistent reuse licensing, or visible governance frameworks.

Participatory platforms such as iNaturalist and OpenStreetMap, while not traditional repositories, illustrate the potential of citizen-generated data when community contribution is embedded by design. Yet, such models remain the exception rather than the norm. Repositories in health and soil-related domains face particular structural and legal constraints that limit openness and reuse.

Overall, the analysis highlights the need for greater harmonisation, transparency, and sustainability in open data systems—ensuring they are not only technically robust, but also ethically governed, socially inclusive, and institutionally aligned with the Missions' policy goals. For initiatives like CROPS, this underscores the importance of identifying scalable models and best practices that enable the upscaling of citizen science, from fragmented pilot efforts to more systemic, policy-integrated contributions across the EU research landscape.

13. References

Bowser, A., Cooper, C., de Sherbinin, A., Wiggins, A., Brenton, P., Chuang, T.R., Faustman, E., Haklay, M., Meloche, M., 2020. Still in Need of Norms: The State of the Data in Citizen Science. Citizen Science: Theory and Practice 5, 18. https://doi.org/DOI: 10.5334/cstp.303

Calyx, C., 2020. Sustaining citizen science beyond an emergency. https://doi.org/10.31235/osf.io/4bf23

Groom, Q., Weatherdon, L., Geijzendorffer, I.R., 2017. Is citizen science an open science in the case of biodiversity observations? Journal of Applied Ecology 54, 612–617. https://doi.org/10.1111/1365-2664.12767

Lush, V., Bastin ,L., Otsu ,K., and Masó, J., 2024. Assessing FAIRness of citizen science data in the context of the Green Deal Data Space. International Journal of Digital Earth 17, 2344587. https://doi.org/10.1080/17538947.2024.2344587

Price-Jones, V., Brown, P., Adriaens, T., Tricarico, E., Farrow, R.A., Cardoso, A.-C., Gervasini, E., Groom, Q., Reyserhove, L., Schade, S., Tsinaraki, C., Marchante, E., 2022. Half a billion eyes on the ground: citizen science contributes to research, policy and management of biological invasions in Europe. ARPHA Preprints 3, e81567. https://doi.org/10.3897/arphapreprints.e81567

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.