D9.2 - Data Management Plan

V1.0

30 June 2023

**Funded by
the European Union**

DELIVERABLE INFORMATION

Deliverable title:	D9.2 – Data Management Plan
Deliverable #:	V1.0
Deliverable Type:	Report
Dissemination level:	Public
WP:	WP9
Responsible beneficiary:	University of Aveiro
Authored by:	Dionisia Laranjeiro (UAveiro), Ana Lillebø (UAveiro)
Reviewed by:	Daniel von Schiller (UB), Chiara Santinelli (CNR)
Due date	30 June 2023
Submission date:	30 June 2023
Deliverable Status:	Submitted
Citation:	RESTORE4Cs. 2023. *Data Management Plan*. Deliverable D9.2 Horizon RESTORE4Cs Project GA ID: 101056782

PROJECT INFORMATION

Title of the project:	RESTORE4Cs
Instrument:	HORIZON-RIA
Contract Number:	101056782
Project coordinator:	Ana Lillebø (University of Aveiro)
Start of the project:	January 2023
Duration:	36 months

Table of Contents

1. Data Summary

The RESTORE4Cs Data Management Plan is an operational document that outlines the main steps and procedures to ensure FAIR data (making data findable, including provisions for metadata, openly accessible, interoperable and increase data re-use), data security and ethical aspects. RESTORE4Cs follows the European Code of Conduct for Research Integrity principles of Reliability, Honesty, Respect and Accountability. Intellectual property rights follow the guidelines on EU trade policy and intellectual property by the European Commission.

Purpose of the data generation or re-use and its relation to the objectives of the project

RESTORE4Cs' will create datasets, tools and scientific publications (e.g., evidence on wetlands Green House Gases (GHG) abatement potential) complying with the FAIR principles, that will be used by different stakeholders, including the research community, policy makers and governmental agencies, environmental managers and the general public. The project aims to increase the transparency, robustness, trustworthiness, and practical usability of the knowledge base on the role of wetland restoration for climate change mitigation that will be used by the scientific community (improved data, improved models, improved conceptual frameworks).

Real-world data, both from in-situ studies at Case Pilot sites and from meta-analysis, will for the first time allow the calibration of response models on the effect of restoration actions on GHG mitigation in wetlands. This will provide the basis, and the tools, to a scientifically based Decision Support System (DSS) for the selection of the best restoration actions and the best sites to be applied by considering the trade-offs. This a key milestone for the implementation of scientific knowledge into the selection of nature-based solutions to mitigate climate change.

Established and reinforced networks, data and information sharing, following FAIR principles, will substantially contribute to key international assessments (e.g., IPCC, IPBES, SOER/EEA).

Information on wetland status and trends, on their capability to mitigate climate change impacts and on the main restoration actions that could improve these capabilities are usually limited, sparse, difficult to obtain and scale and/or wetland type specific. RESTORE4Cs will provide online, user-friendly and integrated tools as unique entry points for wetland practitioners and decision makers regarding the prioritisation of conservation and restoration actions, in relation with their GHG performance as well as impacts on biodiversity and a wide range of ecosystem services.

Stakeholders using the data

Research Community: International academics and researchers will be users of the project data, including processes, outcomes, methodologies, and approaches used, enabling them to start working for the scale up of project's results to all possible wetland (and not only) ecosystems. The data sharing FAIR principles adopted will contribute to strengthen international cooperation and the overall excellence of the European Research Area.

Environmental managers, involved as key actors in the community building process and trained on the DSS (toolbox) developed within the project, will use project data, including processes, outcomes, methodologies, and approaches for decision making (e.g., adaptive management) at different geographical scales.

Policymakers, governmental agencies, EU, national and regional administrative authorities in charge of the implementation of the actions set out in the EU Biodiversity Strategy and the LULUCF regulation, engaged and knowledgeable about the online platform and toolbox for wetland restoration prioritization, status assessment and decision making, will be able to use and benefit from it, supported by the policy-streams implementation roadmaps.

Society: The foreseen actions of engagement and capacity building will promote social awareness on the project outcomes and their impact on the wellbeing of the communities involved in restoration actions, other stakeholders, and citizens. Younger generations will have a pivotal role in long-term awareness.

Types and formats of data the project will generate or re-use

Existing evidence shows the need to scale up wetland assessment tools and information products using available European datasets and big data techniques.

The project involves the development of new information layers and mapping tools (using advanced remote sensing techniques) and the upscaling of existing tools (using big data analytics) to analyse the status of wetland ecosystems and their pressures at broad scales, and to provide a spatially explicit mapping of wetland restoration potential across Europe. These maps (as datasets, documented with standardised metadata) and all derived wetland restoration indicators will be available at the pan-European scale, but also downscaled to national and sub-national levels.

Whenever possible and necessary, RESTORE4Cs will re-use data from open access databases and data available on request from data providers, in particular statistical data. It is foreseen to use small datasets from the following sources – Eurostat European Statistics (the statistical office of the European Union, responsible for publishing high quality statistics and indicators across Europe, allowing comparisons between countries and regions) and institutional repositories of the consortium partners, such as RIA - Repository of the University of Aveiro.

Datasets used by researchers and other stakeholders with interest in Environmental Economics and natural resources, and any other data sources relevant to the project objectives, will be taken into consideration.

RESTORE4Cs will generate and collect a diverse range of data types, including text, audio, numerical and observation data, photos, images, reports, satellite data, programming codes, gaming codes, spreadsheets data files. The data will be collected using various methods such as surveys, interviews, observations, experiments, serious games, machine learning, big data analytics and advanced remote sensing techniques.

Only data paramount to carry out the project activities will be collected, and participants will not be asked to provide personal data unless this is deemed strictly necessary. RESTORE4Cs will use widely accepted formats for data generation, namely: pdf, txt, doc/docx (documents, reports, publications, and policy briefs); xls/xslx (data spreadsheets); cvs (databases); jpg, png (pictures); avi, mp4, wmv (videos).

2. FAIR data

2.1 Making data findable, including provisions for metadata

Persistent identifier

The Findable principle aims to ensure that data can be easily located and identified by both humans and machines. This involves using standardized metadata and persistent identifiers that enable data to be easily discoverable.

The data identifier is a unique identifier that remains unchanged and cannot be duplicated or deleted even if the data to which it refers has been removed. The identifiers should reflect the version of the data. Example of data identifier:

RESTORE4Cs.report.2023.1.v1

Where:

- "RESTORE4Cs" is a reference to the project.

- "Report" refers to the data type. This can be "paper", "code", "num" …

- "2023" is a reference to the year when the data was entered into the RESTORE4Cs common database.

- "1" is the serial number of this data.

- "v1" is the version of the data.

The data identifier must be created by the RESTORE4Cs Data Management Team, and not by the partner/team who produced the data. The partner/team who produced the data should provide comprehensive metadata that describes the data, and this metadata will be accessible by any partner by using the data identifiers.

Naming and versions of a dataset

Researchers will use appropriate file naming conventions and folder structures to ensure that the data is organised and easy to locate. For instance, researchers will use descriptive file names that provide information about the content, format, and date of creation of the data. A version number must be assigned. For a new dataset, the initial version number should be "v1". Any subsequent modification to the dataset, including the addition of new data or changes to

existing data, requires a new version number, such as "v2", "v3", and so on. The use of decimal numbers in versions, such as "v1.1", to denote a modified dataset should be avoided. Additionally, the metadata must include the version number and a description of any modifications made if the dataset has been modified.

Metadata

Metadata is information that describes, contextualises and provides additional information about a particular data or resource to help managing and organising digital resources and making them easier to find and more accessible to users. It can include a wide range of details such as the title, author, creation date, format, location, subject of a digital file or resource, as well as information about how to cite the data. In order to manage the diversity of data types collected in the RESTORE4Cs project, the naming and metadata of each data type must follow the indications of the Data Management Plan.

- Numerical and observation data should be stored in spreadsheets or databases, in .xls or .cvs files, with accompanying metadata such as variable labels, units, location, data generator name and contact, data collection date and time.

- Text data should be stored in text files, with accompanying metadata such as author, date, and document type.

- Audio and video data should be stored in appropriate multimedia formats, with accompanying metadata such as the author(s), recording date, location, and format.

- Satellite data will be stored in GIS formats or TIFF format with metadata such as satellite type, sensor type, author and data acquisition time.

- Programming and gaming codes should be stored in proper script files. The script files should have comment sections that describe the different parts of the code. The metadata of the script files should include proper documentation of the programming languages used, the author(s) and the version.

- Metadata of deposited publications will be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine actionable) and will provide information about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organizations and the grant. Where applicable, the metadata will include persistent identifiers for any research output, or any other tools and instruments needed to validate the conclusions of the publication.

2.2 Making data accessible

The Accessible principle stresses the importance of providing open and easy access to data for all interested parties. This includes ensuring that data is available in formats that can be read by both humans and machines, and that data access is not restricted by legal or technical barriers.

RESTORE4Cs will provide (digital or physical) access to data or other results needed for the validation of the conclusions of scientific publications, to the extent of the legitimate interests or constraints are safeguarded.

Project repository and website

The RESTORE4Cs project is committed to complying with ethical and legal considerations in all the aspects of its research activities. The EU recommends making research data open access, which allows for maximum dissemination and use of research findings. Therefore, the project intends to make the research data generated as part of the project openly accessible to the public.

One task of the project is to create a repository for all data and metadata of interest to the project, including all types of data produced at the scale of the Pilot Sites and those upscaled at national and pan-European levels, taking into account the FAIR principles. This data Repository will be hosted by the University of Salento Data Centre of LifeWatch Italy, and will gather all collected and produced data from the different WPs, including metadata descriptions and a report on the technical baseline.

The RESTORE4Cs website (www.restore4cs.eu), which is already online, supports dissemination and communication purposes, promoting knowledge and data sharing as an open data policy. The RESTORE4Cs' website has a privacy statement that clearly communicates the project's data policy to the public.

Between the Website and the Repository, publications, reports and tools, presentations, datasets, images, video/audio and interactive materials such as lectures will be made available online for download. As the data will be available online, it will be accessible through any web browser application.

To maximise dissemination, partners may publish cumulatively in other trusted repositories, such as the official repositories of their institutions. These must be available free of charge to make research data accessible, following the FAIR principles.

Publications and underlying datasets will be linked through the use of persistent identifiers (DOI versioning). Datasets with dissemination level of "confidential" (non-anonymous datasets) will not be shared due to privacy concerns. Potentially, some datasets might be restricted due to protection for commercial exploitation. If such cases arise during the project, this will be informed in the final version of the Data Management Plan.

Personal data

The project recognizes that personal data may be collected as part of the research activities. Before collecting personal data, researchers will obtain the consent of the ethics committees of their institutions, and they will then obtain the consent of the individuals concerned and inform them of how their data will be used.

The project will take all necessary measures to ensure the confidentiality and security of personal data and will comply with EU regulations regarding personal data protection. To further ensure data protection, the project will maintain a database of personal data for a maximum of 12 months after the project ends. Before the database is destroyed, the project team will contact the individuals who have data in the database to inquire if they wish to continue being part of the database or not. This process ensures that the project team has the most up-to-date information for communication purposes while respecting the rights of the individuals concerned.

2.3 Making data interoperable

The Interoperable principle emphasizes the need for data to be structured in a way that enables easy integration and analysis across different platforms and disciplines. This requires the use of common standards and protocols, as well as the adoption of best practices for data management.

The data management will be based on the most advanced data and metadata standards (e.g., the INSPIRE Directive for spatial information), on controlled vocabularies and quality control procedures. This process requires steps such as the assignment of a unique PID (DOI or permanent URL), the application of standardized procedures and tools specifically developed to make marine data FAIR and to process new data using international standards and formats. The RESTORE4CS team will assess the suitability of specific existing vocabularies to be used for the (meta)data annotation in order to ensure their harmonization and their compliance to FAIR principles through:

- the evaluation and selection of terms (concepts) from the existing vocabularies.

- the development of new terms extending those occurring in existing resources (e.g., LifeWatch thesauri, BODC vocabularies).

- the development of new controlled vocabularies, specific for Coastal Wetlands restoration, by the exploitation of the EcoPortal, i.e., the LifeWatch ERIC repository for ontologies and thesauri with associated essential data curation services.

RESTORE4Cs will observe OpenAIRE guidelines for online interoperability, including OpenAIRE Guidelines for Literature Repositories and the OpenAIRE Guidelines for Data Archives. As the project progresses and data is identified and collected, further information on making data interoperable will be outlined in subsequent versions of the Data Management Plan.

Specifically, information on data and metadata vocabularies, standards or methodology to follow to facilitate interoperability.

2.4 Data re-use

The project will provide open access to peer-reviewed scientific publications relating to the achieved results, ensuring that:

- At the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications.
- Immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g., CC BY-NC, CC BY-ND).
- Information is given via the repository about any research output, or any other tools and instruments needed to validate the conclusions of the scientific publication.
- The authors retain sufficient intellectual property rights to comply with the open access requirements.
- Scientific articles will be published in full open access venues for peer-reviewed scientific publications, favouring open access platforms such as Open Research Europe.

To maximize impact, it is also recommended to publish research papers or reports that document the methodology used to collect and handle the data, as well as any significant findings or conclusions drawn from the data. These publications should cite the DOIs of the relevant data repositories (e.g., Zenodo and others recommended by the CE).

Open access publications are usually released with the associated research data available on supplementary material. Whenever an embargo period is needed, the publication and the related data will be made available after the embargo period has expired. A common time span for embargoing is between 6 and 24 months. Data published in open access repositories will be available for the lifetime of the repositories, accordingly to their general policies.

Personal data should not be redistributed under any conditions.

Data citation

All users who access and use the data generated or collected during the project are required to properly cite the data according to the guidelines provided in the metadata.

3. Other research outputs

For other research outputs, such as digital (e.g., software, workflows, protocols, models, etc.) or physical (e.g., new materials, reagents, samples, etc.), the guidelines and provisions already defined for data will be applied.

These policies and provisions aim to promote transparency, collaboration, and maximum data sharing and re-use while protecting the intellectual property rights of the data producers and ensuring proper citation and acknowledgment of the data users.

4. Allocation of resources

The research outputs of the RESTORE4Cs project will be Open Access, i.e., free internet access to and use of publicly funded scientific publications and data following the FAIR principles. The RESTORE4Cs foresees dedicated budget for Open Access publications, distributed by the partners, and associated to Copernicus DIAS access (WP6 platform).

RESTORE4Cs has a Data Management Team responsible for:

- Preparing and developing the necessary infrastructure to implement the Data Management Plan.

- Review and notify researchers and teams of any missing metadata.

- Update the Data Management Plan as requested by the RESTORE4Cs Consortium or the Coordinator.

- Work with all partners, the RESTORE4Cs Consortium and the Coordinator to ensure the destruction of personal data after 12 months from the end of the project.

- Work with all partners, the RESTORE4Cs Consortium and the Coordinator to increase the visibility and impact of the RESTORE4Cs data.

5. Data security

The following section provides guidance on how to store and back up various types of data. To simplify matters, data storage is divided into three levels based on the stage at which the data is being handled: 1) during data collection/creation, 2) temporary storage for further handling by the same or other teams, and 3) long-term storage.

Storage during the data collection/creation stage

Researchers and different teams can store data on their personal devices (e.g., PCs, cameras) during data collection and handling stages. The researchers should make sure that:

- Data should be stored properly.

- Access to the data should be provided to others handling the data at this early stage through a common platform (e.g., MS Teams or any other platforms).

- Data should be backed up on an external drive and on a cloud service, if possible.

- The researcher should prepare the metadata according to the format defined.

- The metadata should include a description of the methods used to collect the data and the programs that may be used to handle the data.

Temporary storage for further handling by the same or other teams

At this stage, the data are parked for 1) further handling by the researchers or by other RESTORE4Cs partners and/or 2) waiting for publications (e.g., reports, research papers or data repositories). The researchers and the RESTORE4Cs' data management team should make sure that:

- Data should be stored properly.

- The researcher should keep at least a copy of the data on an external drive.

- The RESTORE4Cs data management team should keep at least three copies of the data (a copy that will be made searchable and accessible to all the RESTORE4Cs partners through a common platform, a second copy that should be saved on a cloud service but inaccessible, and a third version that is saved on an external drive).

- All the entered data should have a unique data identifier.

- The data identifiers for all three versions of the data should be identical and the data itself should be completely identical (100% similar). Any modification of the data requires creating a new data version.

- The metadata is completed in a proper format and uploaded and is therefore searchable and accessible by other teams.

Long-term storage

Data will be safely stored, according to the best practices, in institutional repositories and/or certified ones, allowing long-term availability. Data retention and preservation is further presented in section 2.2.

6. Ethics

RESTORE4Cs coordination is supported by the Ethics and Deontology Council (EDC) of the University of Aveiro, which is responsible for providing the appropriate guidelines for the establishment and consolidation of a policy that safeguards ethical and deontological

principles, by issuing opinions, when requested or proposing, on its own initiative, the adoption of codes of conduct. Specific tasks may be supported by the Ethics Departments of other partners involved.

Some research topics involve transdisciplinary approaches that imply gathering information and public discussion sessions with adults capable to understand and sign the informed consent (e.g., focus groups, workshops, world cafe). The participation in these studies is always voluntary and, in line with national and EU regulations on participatory processes, participants are requested to sign an Informed Consent Form.

The participation of children and/or adults unable to give informed consent is not foreseen, yet if deem necessary it will be voluntary and, in line with national and EU regulations on participatory processes. Authorization for children's participation will be given by the legal guardian at the stage of participants recruitment. All the studies involving children and/or adults unable to give informed consent will be previously submitted to the EDC of University of Aveiro.

Concerning protection of personal data, whenever required, studies will be previously submitted to the Portuguese Data Protection Authority. All participants are volunteers and Informed Consent will be obtained from each participant. The anonymisation of the participants will be always assured (e.g., questionnaires), in line with national and EU regulations on data protection. The access to this data is restricted to the researchers directly associated with the development of the study, and files associated with this information are password protected.

7. Other issues

Currently, the project does not use any data management procedures other than those described in this Data Management Policy.

The RESTORE4Cs Data Management Plan is a living document that will be continuously updated during the course of the project. Day-to-day data management will be carried out through ongoing collaboration between the Data Management Team, the Consortium Partners and the Coordinator. A revised and updated Data Management Plan will be submitted in month 33 (D9.3).