PNRR ECCSELLENT

Data policy specifications

D. 5.2

(version 1.0)

Document Information Summary

Date	11/09/24
Document title	ECCSELLENT WP 5 Data Policy Specification
ID	5.2
Leading Partner	OGS
Main Author(s), Organization	A.Viola, P. Diviacco, OGS
Work package / Activity	5
Involved O.U.	UO 1 OGS - Geophysics
Reviewer(s)	
Approved by	

Keywords	Data policy, FAIR, Data management

Target audiences	Project Partnership	
Deliverable nature	O Report	O Other (specify)……
Dissemination level	O Public	O Restricted to participants

Revision history

Version	Date	Description
2.0	13/05/2024	Draft

Finanziato dall'Unione europea
NextGenerationEU

PNRR "ECCSELLENT" - Missione 4, "Istruzione e Ricerca" - Componente 2, "Dalla ricerca all'impresa" - Linea di investimento 3.1, "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione" Finanziato dall'Unione europea NextGenerationEU
CUP …

TABLE OF CONTENT

Finanziato dall'Unione europea NextGenerationEU

PNRR “ECCSELLENT” - Missione 4, “Istruzione e Ricerca” - Componente 2, “Dalla ricerca all’impresa” - Linea di investimento 3.1, “Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione” Finanziato dall'Unione europea NextGenerationEU
CUP …

LIST of ACRONYMS

DOI Digital Object Identifier

FAIR Findable, Accessible, Interoperable, Reusable

I.O. Intermediate Objective

O.U. Operating Unit

SUMMARY

This document (data policy) provides guidelines for the handling of data within the ECCSELLENT project. It ensures that data is handled in a consistent and transparent manner. The ECCSELLENT project aims to promote the development and internationalization of our country's research in the entire CCUS chain: CO2 capture, utilization, transport, and storage. In this context, the integration and networking of data is essential. When data providers agree to share, they must be assured that their data will be properly handled, disseminated, and recognized by all stakeholders according to similar principles and rules.

Finanziato dall'Unione europea
NextGenerationEU

PNRR "ECCSELLENT" - Missione 4, "Istruzione e Ricerca" - Componente 2, "Dalla ricerca all'impresa" - Linea di investimento 3.1, "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione" Finanziato dall'Unione europea NextGenerationEU
CUP …

PREAMBLE

The policy contains guidelines on how ECCSELLENT handles data. This ensures that data is handled in a uniform and transparent manner. The ECCSELLENT project aims to promote the development and internationalization of our country's research in the entire CCUS chain (capture, use, transport and storage of CO2). In this context, the integration and networking of data is essential. When data providers agree to share data, they must be assured that their data will be properly handled, disseminated and recognized and that similar principles and rules apply to all parties involved.

Partners and facilities involved in the project:

The partners, coordinated by the National Institute of Oceanography and Experimental Geophysics - OGS, are the University of Bologna, Politecnico di Milano (PoliMI), the National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA) and the National Research Council (CNR).

RECOGNISING:

The general objectives of the National Recovery and Resilience Plan (PNRR) (https://www.governo.it/sites/governo.it/files/PNRR_0.pdf) and the specific objectives of the PNRR, MISSION 4 "Education and Research" - COMPONENT 2 "From research to business", INVESTMENT LINE 3.1.1 "Fund for construction of an integrated system of research and innovation infrastructures" (Research Infrastructure). The objectives are:

- strengthen research and encourage the dissemination of information models for basic and applied research conducted in synergy between universities and businesses,
- support processes for innovation and technology transfer,
- strengthen research infrastructures, capital and skills to support innovation.

The Infrastructure for Spatial Information in the European Community (INSPIRE) (https://knowledge-base.inspire.ec.europa.eu/index_en), which addresses spatial data themes needed for environmental applications, aims to provide relevant, harmonized and high-quality spatial data to support the formulation, implementation, monitoring and evaluation of policies and activities that have a direct or indirect impact on the environment.

The INSPIRE Directive establishes harmonized conditions for access to spatial data sets and services and facilitates the sharing of spatial data sets and services by public authorities within and between Member States and by Community institutions and bodies.

ECCSELLENT Data Policy

1 SUBJECT MATTER AND SCOPE

The purpose of this Data Policy is to outline the requirements relating to data sharing, access, retention, and attribution in order to facilitate the widespread use and re-use of data and information. This policy applies to data collected, acquired, processed, and disseminated within the ECCSELLENT project. Thus, this policy applies both to data owned by the OGS and to data owned by others. Data, metadata, and products should be shared in an accessible, reproducible, interoperable, free and open manner with minimal delays and restrictions. Such data sharing, both real-time and delayed, facilitates scientific research and innovation.
Data owners or producers will support, promote, and enable full, free, and open access to specific types of data. Data will be made available to the project consortium or published under an open-source license (e.g., Zenodo repository) unless there are legitimate reasons to the contrary relating to intellectual property rights and confidentiality.
This approach ensures that data can be freely accessed, used, modified, and shared according to the principle "as open as possible, as closed as necessary."
Third parties must always comply with legal, ethical, and regulatory frameworks, including appropriate acknowledgment, when using the data. This policy allows for the citation and licensing of data by public access while recognizing the need for the results' owner to preserve confidentiality prior to scientific publication and intellectual property protection.

2 OBJECTIVES

The objectives of the data policy are to support, promote and enable:

- the continuous availability of up-to-date data and the maintenance of long-term series of observations,
- the wider use, reuse and recombination of data from different sources in different frameworks and media than those for which they were originally commissioned,
- full, free and open access to all types of data wherever possible ,
- the protection of integrity, transparency and traceability of data, recognition of data providers and their intellectual property rights,
- compliance with relevant legislation and government guidelines for data management and dissemination,
- implementation of INSPIRE and F.A.I.R. data principles,
- to identify data uniquely and permanently through the use of Digital Objects Identifiers (DOI),

Finanziato dall'Unione europea NextGenerationEU

PNRR “ECCSELLENT” - Missione 4, “Istruzione e Ricerca” - Componente 2, “Dalla ricerca all’impresa” - Linea di investimento 3.1, “Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione” Finanziato dall'Unione europea NextGenerationEU
CUP …

- to build metadata models in order to be compliant with Open Geospatial Consortium (OGC) and ISO standards such as 19115, SensorML, O&M or Dublic Core,
- publication of relevant metadata

3 ACCESS TO AND REDISTRIBUTION OF DATA

Access to data includes both technical access and the policy that regulates access.

Data and associated metadata should be made available with minimal restrictions on use unless there are valid reasons to restrict access. Legitimate reasons for restricting access to and reuse of data and associated metadata include privacy and confidentiality, protection of rare or endangered species, and national security.

Data will be made available through discovery, display and, where possible, through download services that conform to established standards of ISO, OGC, INSPIRE and other relevant standardization bodies. The OGS will aim to provide meta-information for all data. Unless otherwise stated, the dataset is distributed under the Creative Commons Attribution License (CC-BY) standard license (https://creativecommons.org/share-your-work/cclicenses) or a similar license.

The project's discovery and data access system will be connected to the main data system through dedicated portals. Queries will be handled via a dispatcher front-end portal, where data requests will be redirected to a domain specific data portal. This mirrors the detachment between the different metadata levels and guarantees the modularity of the system. When a new type of data will enter the data management system, a specific new module will be developed leaving the other modules untouched. The isolation of domains will allow rapid new developments, less downtime, and less trouble with other communities of practices. Domain specific data portals can also be already existing data portals outside the project if they are already consolidated and reference data hubs for a community of practice or a research area.

Geospatial data will be compliant with the INSPIRE directive and available via compliant network services such as OGC WMS and WFS.

Harvesting of metadata will be enabled by OAI-PMH services, so that other data sharing initiatives will be able to redirect users to ECCSELLENT services.

4 CONDITIONS OF USE

The data should be licensed under a minimally restrictive and voluntary common-use license that grants permission under specific conditions, ensures proper attribution, and allows others to copy, distribute, and use the data. The Creative Commons Attribution License (CC-BY) is recommended, but not mandatory.

5 F.A.I.R. Principles

To support knowledge discovery and innovation by both humans and machines, data should conform as much as possible to the F.A.I.R. (Findable, Accessible, Interoperable and Reusable) guiding principles.

Semantic technologies such as controlled vocabularies and ontologies are used to enable cross-disciplinary integration of metadata.

Among the various technical details of the implementation of the F.A.I.R. concept, the use of Digital Objects Identifiers (DOI) is a very important segment. A DOI separates the actual storage location of the data from the identifiers so that the link can be managed in due time and users always have up-to-date options for accessing the data.

6 QUALITY

The main responsibility for the quality of the data they produce and disseminate remains with the data providers. For the data produced by the OGS, the OGS shall strive to publish quality metadata that includes information on transparency, accuracy, relevance, timeliness, consistency, and comparability where appropriate.

7 DATA REPOSITORIES

Repositories should comply as far as possible with the F.A.I.R. guiding principles and meet the requirements for data access, preservation, resource stability and licenses.

Repositories must:

- Ensure the long-term maintenance and preservation of datasets;
- Provide stable identifiers for datasets, preferably in the form of a DOI;
- Provide public access to data as far as possible;
- Support open licenses.

The use among partners of already existing repositories such as, for example, Pangaea (https://www.pangaea.de/) or Zenodo (https://zenodo.org/), where data products can be made available in an open data perspective, is considered if they are a common practice in a domain community and if they are compliant with F.A.I.R. and INSPIRE principles. OGS maintains a F.A.I.R. and INSPIRE compliant data management system called SNAP (https://snap.ogs.it/) where specific data types such as seismics, borehole and a generic data lake repository are handled.

8 SECURE LONG-TERM DATA ARCHIVES

The archiving of data and metadata is guaranteed in accordance with the provisions of the National Recovery and Resilience Plan (PNRR).

9 DATA PRODUCTS

Each partner of the project specifies in the following table the data product type that will need to be handled within the project and the possible external data repository used.

Partner	Data product type	Data handled within Eccsellent	Possible external repository	DOI minting organisation
OGS	Geophysical data	X		OGS
OGS	Marine Biology	X		OGS
OGS	Oceanography	X		OGS
PoliMi	Process simulations and tests or laboratory analyses data results (Excel, Matlab, Python, Aspen Plus or other formats produced by specific commercial software)	X	OneDrive private repository shared with few members of the team	
CNR-ITAE	Standardized (internal .csv files.			
CNR-STEMS	Standardized (internal) .csv, .jpg, tiff			
ENEA	SESR, DRM, SEG tests data results (.csv, .xlsx, .pdf, .txt, opju)	X		
ENEA	XRD, SEM, BET, TGA, XPS, etc. results (proprietary files format)	X		
ENEA	Process simulations (.xlsx, .mat)	X		
ENEA	CPFD model (.stl)	X		
UniBO	No stand., Excel files			

10 FUTURE DATA POLICY REVIEW

Technological change in information and communication technologies has innovative effects on the collection, processing and use of data. This data policy aims to explore and exploit these

opportunities. In order to take full advantage of these developments, this policy will be reviewed at regular intervals.

11 DEFINITIONS

- Complete, open and freely available means complete, non-discriminatory and free of charge.
- Data are defined as elements or records regardless of whether they were obtained directly with instruments or from legacy or vintage records, whether they are made available from processing steps or originate from other sources and initiatives.
- Metadata is 'data about data' that describes the content, quality, condition and other characteristics of data that enable its inventory, discovery, evaluation or use.
- Digital Objects Identifiers (DOI) are strings of numbers, letters and symbols that can be used to uniquely identify a research artifact and provide it with a permanent web address (URL).
- Products and services are all information resulting from the transformation or processing of data in the form of assessments, web services, images, diagrams, texts or files containing essential know-how. Usually associated with added value.
- Redistribution means transfer to a third party other than the originator of the data and products.
- Reuse means the use of data and documents held by public sector bodies by natural or legal persons for commercial or non-commercial purposes other than the original purpose of the public task for which the data and documents were created.