Hymantovalley

D.1.3 DATA MANAGEMENT PLAN

PROJECT	
Project number:	1011115147
Project acronym:	HYMANTOVALLEY
Project name:	Hydrogen Valley in the Province of Mantova

DATA MANAGEMENT PLAN	
Date:	12/04/2024
Version:	1.4

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

The Hymantovalley project could re-use data from previous project and could exchange data with other current project about green hydrogen and logistic, developed in Mantova in order to complete a whole green hydrogen value chain:

- WIN-IT - Works for Implementing the Navigation in Northern Italy
- CRISTAL - Climate resilient and enviromentally sustainable transport infrastructure, with a focus on inland waterways
- K-WAY - Improving Key Italian inland WAterwaY infrastructures by means of punctual interventions and general investment action
- MANTHOVA: Mantua Hydrogen Valley
- HYPER MANTOVA - Hydrogen High Pressure Efficient Renewing At Mantova Facility: Innovazione Ed Efficientamento Della Filiera Di Distribuzione dell'Idrogeno Compresso
- PNRR green hydrogen production from renewable energy sources in decommisioned industrial areas
- PNRR HyPER Mantova - Hydrogen High Pressure Efficient Renewing at Mantova Facility
- PNRR loal public transport through electric and hydrogen buses and related facilities
- Proposte per la strategia italiana dell'idrogeno e la competitività economica (https://www.ambrosetti.eu/en/site/get-media/?type=doc&id=20020&doc_player=1)
- Hydrogen Valleys (https://www.clean-hydrogen.europa.eu/hydrogen-valleys-0_en)
- Hylaw online database (https://www.hylaw.eu/)
- Mission Innovation Hydrogen Valley Platform (https://h2v.eu/)
- sHYpS: Sustainable HYdrogen powered Shipping (https://www.shyps.eu/)

What types and formats of data will the project generate or re-use?

Hymantovalley will re-use, generate and collect:

- Technical, social, economic and environmental data and metadata from previous projects, parallel/contemporary projects, hydrogen and logistic networks, existing european and national databases, field experiments and demonstration platforms.
- Knowledge, experiences and insights based on stakeholder interactions.
- Personal data, records and images.

All personal data will be collected and processed in accordance with General Data Protection Regulations (GDPR - Regulation (EU) 2016/679), respecting the principles of the Regulation, in particular:

- Lawfulness, fairness, and transparency;
- Purpose limitation;
- Data minimisation;

- Accuracy;
- Storage limitation;
- Integrity and confidentiality;
- Accountability.

Data will also be used in physical models and relational databases.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

Data will be generated and re-use in order to:

- Testing and demonstrating the green hydrogen value chain , in particular:
    - Technical and economic advantages of an electric accumulator between PV production and the electrolyzer. Sizing evaluation;
    - High pressure hydrogen storage;
    - Mobility and logistic application
    - H2 to heat and power application.
- Defining business models and market analysis;
- Planning a hydrogen research hub suited to the needs of the area, including companies, universities, schools, public authorities, citizens.
- Developing a complete LCA and S-LCA analysis.
- Implementing an effective communication plan, respectful of the rules for processing public and private data and intellectual property

What is the expected size of the data that you intend to generate or re-use?

The expected size of data is about 40 files of an average size of about 5 MB for a Total Data Size (TDS) of 200 MB.

What is the origin/provenance of the data, either generated or re-used?

Data can be provided by each project partner experience/project activities, from public database, from stakeholders through interviews or surveys.

To whom might your data be useful ('data utility'), outside your project?

The generated and collated data will be useful for different kind of stakeholders as:

- Renewable energy companies/designers/scientists/researchers
- Hydrogen companies (producers, distributors, storage experts)
- Universities and research centers
- Public bodies/territorial planners/Public authorities

- Logistic companies

- Final users including companies involved in hard to abate industrial processes, small and medium enterprises.

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Will metadata be offered in such a way that it can be harvested and indexed?

Metadata will be collected and integrated into new or existing databases to enhance accessibility and analysis. The data will be identifiable and locatable using a standard identification system, such as persistent and unique identifiers like Digital Object Identifiers.

No sensitive personal data will be collected, but some personal data, including contact details (such as name, address, email address, and telephone number), will be collected.

This personal data will be stored separately from the project data, with the only connection being a code accessible only to the lead researcher in each country where the data is collected. Therefore, according to GDPR, the project data will not be fully anonymized until the code is destroyed. All participants will be informed of this during the informed consent process.

To avoid misleading information, participants will be told that the research data will only be linked to their personal data by a code, kept securely and separate from the research data, and accessible only to the lead researcher.

They will also be informed of the duration for which their personal details will be retained, and when and if the code/link will be destroyed.

All metadata related to new data generated by the project will be collected on laptops or handheld devices and then uploaded to a central storage area. For activities involving audio or video recordings, the data will be transcribed, and the recordings will be erased or destroyed afterward.

No video recordings will be used for formal data collection, but separate video recordings may be used for dissemination activities. In such cases, individual consent will be obtained.

2.2. Making data accessible

Repository:

Will the data be deposited in a trusted repository?

The collected data restricted to the members of the project will be deposited in the Hymantovalley Teams repository:

https://agenziaagire.sharepoint.com/:f:/s/hymantovalley/EjkNlmW9lxRHhWHenParQYQB49qu5VMuUFq2VdpZYpSlzQ?e=bDuhyN

The project will use file naming conventions that are consistent, concise, meaningful and findable.

The most recent version of each file will be retained on this project internal communication platform along with

relevant older versions.

Public data will be available in the deliverables area of the Hymantovalley website:

https://hymantovalley.agirenet.it/en/

All the project deliverables will be also published on Sygma (System for grant agreement management) through the EU Participant Portal (EU Funding & Tenders Portal).

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Will the data be accessible through a free and standardized access protocol?

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

How will the identity of the person accessing the data be ascertained?

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

The following artefact naming convention will be used:

(XX).(DocumentName).(ProjectName).(yyyy-mm-dd).v(x.x), where:

- (XX) (two numerical characters) unique artefact number within the folder indicating the artefact sequence.

- v(x.x) indicates the artefact version. Version numbers like "0.x" mean that the document hasn't been approved yet; minor changes will be reflected in the decimal (revisions number) and major changes (formal reviews) in the version number.

File names should not exceed 50 characters.

Version control tables will be used for important documents, including version number, date, author, date and whether the document is ‘clean’ or includes ‘tracked changes’ changes.

The Hymantovalley Grant Agreement includes two different levels of dissemination for the project deliverables:

- PU -Public (fully open);

- Sensitive — limited under the conditions of the Grant Agreement.

Deliverables and related data classified as Public will be available as open data for download in the deliverable area of the Hymantovalley website.

Deliverable classified as sensitive will be available as closed data (or on request data) for download only for project partners (following the rules of the Grant Agreement and Consortium Agreement and considering intellectual propriety management).

2.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data

interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Will your data include qualified references1 to other data (e.g. other data from your project, or datasets from previous research)?

The data generated by the project will be interoperable, enabling data exchange and reuse among researchers, institutions, organizations, and countries.

Hymantovalley will follow standards for formats, ensuring compatibility with available (open) software applications as much as possible. This approach will particularly support the integration and re-combination of different datasets from various sources.

2.4. Increase data re-use

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Will the data produced in the project be useable by third parties, in particular after the end of the project?

All open data produced and/or used in this project will be made available for use by third parties. The data will be released under an appropriate open license (such as Creative Commons Attribution or similar), ensuring that it can be freely accessed, reused, modified, and shared, subject to proper attribution.

Will the provenance of the data be thoroughly documented using the appropriate standards?

The provenance of the data will be thoroughly documented to ensure transparency and reproducibility.

This will be achieved by:

- Using standard metadata schemas (such as Dublin Core, DataCite Metadata Schema) to record the origin, lineage, and changes to the data.
- Implementing version control systems (such as Git) to track modifications and updates to datasets.
- Providing comprehensive data dictionaries and codebooks describing the structure, content, and meaning of the data.o track modifications and updates to datasets.
- Providing comprehensive data dictionaries and codebooks describing the structure, content, and meaning of the data.

Describe all relevant data quality assurance processes.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

Beyond the FAIR (Findable, Accessible, Interoperable, Reusable) principles for data, the DMP will also address other research outputs, such as software, algorithms, and publications.

---

[1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

- Findable: Research outputs will be indexed in relevant repositories and databases
- Accessible: Outputs will be made openly accessible via institutional repositories or subject-specific archives.
- Interoperable: Standards and formats that facilitate data interoperability will be adopted (such as CVS or XML).
- Reusable: Outputs will be accompanied by clear licenses and detailed documentation to enable reuse.

Resources for data management will be adequately allocated within the project budget.

This includes:

- the personnel for data management tasks, including data stewards and IT support.
- Infrastructure costs for data storage, backup, and preservation.
- Tools and services for data processing, analysis, and sharing.

Data security measures will be in place to protect sensitive and confidential information:

Access Controls: Role-based access controls will be implemented to restrict data access to authorized personnel only.

Encryption: Sensitive data will be encrypted during storage and transmission.??

Backup: Regular backups will be performed to prevent data loss.

Compliance: Data handling procedures will comply with relevant regulations and standards, such as GDPR.

Ethical considerations will be thoroughly addressed:

- Informed Consent: Participants' informed consent will be obtained, ensuring they understand the purpose, use, and sharing of their data.
- Anonymization: Personal data will be anonymized or pseudonymized to protect individuals' privacy.
- Ethics Review: The project will undergo ethics review and approval by appropriate ethics committees to ensure compliance with ethical standards.

3. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

Who will be responsible for data management in your project?

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Costs for making data FAIR will be covered by the organisations involved in the gathering and analysis of data, and therefore ultimately in the publication of related research papers. Costs related to open access to research data are eligible as part of the grant.

Data management will be the responsibility of all *Hymantovalley* participants involved in data collation, collection, transfer and storage. Task leaders, WP leaders and the coordinator will be responsible for the

development of data management protocols and it is the responsibility of other participants to implement these protocols.

Data security will be assured by the existing security protocols of each institution.

All deliverables, including notes, interviews, photographs and videos, will be stored on the participating university's (UNIMORE, POLIMI) research centre (FBK, CNR) and Agirenet secure drive storage.

In the event of an internet connection being unavailable at a particular location or for ease of analysis, these materials may be temporarily stored on individual password-protected personal computers or laptops.

Performance data collected direct from machinery will be stored and processed on-site and held by the company from where the data was collected. It may also be stored temporarily on industries password protected and Agirenet laptops, on university servers and on compliant and secure systems.

4. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Will the data be safely stored in trusted repositories for long term preservation and curation?

The Work package leaders together with the project coordinator are parts of the Project Management Board that is also in charge of ethics advisory committee and will ensure rigorous application of ethical standards, and data privacy standards, and compliance with the General Data Protection Regulation (Regulation (EU) 2016/679), especially in research where human participants are involved. Membership will comprise representatives of Università di Modena e Reggio Emilia (UNIMORE) and Fondazione Bruno Kessler (FBK).

The committee will be tasked with ensuring ethical compliance with the GDPR, through establishing best practice, monitoring ethical procedures, including that associated with human participants, overseeing the collation and archiving of ethical documentation. The committee will meet every six months in person at plenary meetings, and remotely in the intermediary period between meetings. If needed, ad hoc virtual meetings will be called by the chair to deal with specific ethical issues arising.

5. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

The Work package leaders together with the project coordinator are parts of the Project Management Board that is also in charge of ethics advisory committee and will ensure rigorous application of ethical standards, and data privacy standards, and compliance with the General Data Protection Regulation (Regulation (EU) 2016/679), especially in research where human participants are involved. Membership will comprise representatives of Università di Modena e Reggio Emilia (UNIMORE) and Fondazione Bruno Kessler (FBK).

The committee will be tasked with ensuring ethical compliance with the GDPR, through establishing best practice, monitoring ethical procedures, including that associated with human participants, overseeing the collation and archiving of ethical documentation. The committee will meet every six months in person at plenary meetings, and remotely in the intermediary period between meetings. If needed, ad hoc virtual meetings will be called by the chair to deal with specific ethical issues arising.

6. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

SUMMARY TABLE - FAIR Data Management at a glance: issues covered in the HYMANTOVALLET Data Management Plan

Data Management Plan Component	Issues addressed
1. Data summary	- Purpose of the data collection/generation - Relation to the objectives of the project - The types and formats of data generated/collected - How existing data will be used - Origin of the data - Data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	- Discoverability of data (metadata provision) - Identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? - Naming conventions used - Approach for clear versioning - Standards for metadata creation.
2.2 Making data openly accessible	- Data to be made openly available - How the data will be made available - Methods or software tools needed to access data - Where the data and associated metadata, documentation and code are deposited - How access will be provided in case there are any restrictions
2.3. Making data interoperable	- Interoperability of the data. - Data and metadata vocabularies, standards or methodologies to be used. - Use of standard vocabulary for all data types to allow inter-disciplinary interoperability
2.4. Increase data re-use (through clarifying licences)	- How the data will be licenced. - When the data will be made available for re-use.

- Whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project.

3. Allocation of resources

- Estimated costs for making data FAIR
- Responsibilities for data management
- Costs and potential value of long term preservation

4. Data security

- Data recovery, secure storage and transfer of sensitive data

5. Ethical aspects

- See ethics section of DoA and ethics deliverables.

6. Other

- Other national/funder/sectorial/departmental procedures for data management.

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.1	20.05.2024	Initial version (new MFF).