

D1.2. Data Management Plan

Prepared By:

**Antonello Fiorucci
Maddalena Lukasik**



Deliverable Information Sheet

Version	
Grant Agreement Number	101112055
Project Acronym	HYScale
Project Title	HYSCALE – ECONOMIC GREEN HYDROGEN PRODUCTION AT SCALE VIA A NOVEL, CRITICAL RAW MATERIAL FREE, HIGHLY EFFICIENT AND LOW-CAPEX ADVANCED ALKALINE MEMBRANE WATER ELECTROLYSIS TECHNOLOGY
Project Call	HORIZON-JTI-CLEANH2-2022-2
Project Duration	48 months
Deliverable Number	D1.2
Deliverable Title	Data Management Plan
Deliverable Type	DMP — Data Management Plan
Deliverable Dissemination Level	PU - Public
Work Package	WP1
Lead Partner	META
Authors	Antonello Fiorucci; Maddalena Lukasik
Contributing Partners	
Reviewers	
Official Due Date	30-11-2023
Delivery Date	

History of changes

Version	Date	Description	Author(s)
0.1		First draft submitted to partners for validation	Antonello Fiorucci; Maddalena Lukasik

List of Acronyms

DPM	Data Management Plan
DoA	Description of the Action
GDPR	General Data Protection Regulation
FAIR	Findable, Accessible, Interoperable, And Re-Useable
PID	Persistent Identifiers
WP	Work Package

Keywords list

- Hydrogen
- Electrolysis
- Anion Exchange membrane
- Stack development
- Materials development
- Renewable energy storage
- Life Cycle Assessment
- Techno Economic Analysis
- System Engineering

Disclaimer

Co-funded by the European Union. Views and opinions expressed in this document are however those of the author(s) only and do not necessarily reflect those of the European Union or the Clean Hydrogen Partnership. Neither the European Union nor the granting authority can be held responsible for them.

1. Table of Contents

Deliverable Information Sheet	2
History of changes.....	3
List of Acronyms	4
Keywords list	5
Disclaimer	5
1. Table of Contents	6
2. Executive summary	7
3. Data Origin and objectives	8
4. FAIR Data	9
3.1 Making data findable, including provisions for metadata	9
3.2 Making data openly accessible.....	10
3.3 Making data interoperable	11
3.4 Making data reusable	11
5. Data security	12
6. Private data collected via the HYScale events	12
7. Ethics	13

2. Executive summary

Scheduled Data Management Plan (DMP) Updates

The DMP is a document that evolves during the lifespan of the project and registers all relevant changes in the life cycle of all the research datasets of HYScale. Updated versions of the DMP have already been planned (once per year). Moreover, this document will be updated whenever important changes in the data or the data management policy occur.

The Data Management Plan (DMP)

This document constitutes **D1.2 – Data Management Plan**, elaborated in the framework of **WP1 – Project Management**, provides details regarding all the research data collected and generated within HYScale. It explains how research data are handled, organized, licensed, and made openly available to the public and how they will be preserved after the project is completed. The DMP also provides motivations when versions or parts of the project research data cannot be openly shared on account of third-party copyright issues, confidentiality, or personal data protection requirements or when open dissemination could jeopardize the project's achievements. This DMP reflects the current state of the art of HYScale. However, the details and the final number of the project datasets may vary during the research. The variations will be recorded in updated versions of this DMP.

3. Data Origin and objectives

To support the large-scale deployment of clean hydrogen technologies by 2030, the European Union launched the **European Clean Hydrogen Alliance**. In the framework of this partnership has been funded HYScale, a project aiming at upscaling an efficient, durable, sustainable and cost-effective advanced alkaline membrane water electrolysis technology capable of producing economic green hydrogen. Throughout its operation, HYScale will collect and generate different types of data. The current document, which constitutes the first version of HYScale Data Management Plan, aims to establish standards for handling the collected and/or generated data during the lifetime of the project.

The project results will be validated and demonstrated at partners' premises. The project will primarily generate new datasets and will reuse old ones deriving from other EU funded projects.

The following activities will originate new data:

WP	2
Chosen early epository	Local network, Teamsite
Partners involved	CENMAT , DLR, CNR, BEK, CEA
Contact Person	Julien Fage
Data Contents	Cell test, Material analysis, Ex-situ analysis
Data Format	image data (tiff) for ex situ analysis, Specific formats for raw data or .txt files if possible
Accessibility	Only project internal
Possible publications	Cell component benchmarking
WP	3
Chosen early epository	Local network, Sharepoint
Partners involved	CNR, DLR, CEA, PPC
Contact Person	Nicola Briguglio, Antonino Aricò
Data Contents	Stack design, modeling and simulation data, material properties, Ex-situ analysis, electrochemical characterization of components.
Data Format	CAD formats (stack design), image data (tiff) for ex situ analysis and specific formats for raw data or .txt files if possible, excel
Accessibility	Only project internal
Possible publications	Stack design, simulation of the components, Ex-situ analysis/ electrochemical characterization
WP	4
Chosen early epository	Local network, Sharepoint
Partners involved	DLR, CNR, HYG, BEK, CENMAT, CEA

Contact Person	Tobias Morawietz
Data Contents	Stack design, modeling and simulation data, material properties, Ex-situ analysis
Data Format	CAD formats (stack design), image data (tiff) for ex situ analysis and specific formats for raw data or .txt files if possible
Accessibility	Only project internal
Possible publications	Stack design, simulation of the components, Ex-situ analysis
WP	5
Chosen early epository	Local network, Teamsite
Partners involved	HYG, CENMAT, DLR, CNR, DEI
Contact Person	Hans ten Dam
Data Contents	Design, manufacture and assembly of the 100 kW system. Demonstration of the HyScale electrolyzer.
Data Format	*.xls *.doc format
Accessibility	Only project internal
Possible publications	Demonstration results of $> 2A/cm^2$ < 50 kWh/kg H ₂ and a maximum degradation of 0.9%/1000h
WP	6
Chosen early epository	Local network, Sharepoint
Partners involved	All
Contact Person	Mitja Mori
Contents	Technical parameters, Life cycle inventory data (component level, system level), economic parameters.
Data Format	*.xls format
Accessibility	Only project internal
Possible publications	LCA results of (i) component manufacturing, (ii) hydrogen production on system level, (iii) LCOE of produced hydrogen (system level)

4. FAIR Data

This DMP follows the EU guidelines¹ and describes the data management procedures according to the FAIR principles². The acronym FAIR identifies the main features that the project research data must have to be findable, accessible, interoperable, and re-useable, allowing thus for maximum knowledge circulation and return of investment.

3.1 Making data findable, including provisions for metadata

To improve the findability of research data produced during the HYScale project, dataset, beyond locally, will be also deposited in trusted data repositories if and when appropriate.

¹ Guidelines on FAIR Data Management in Horizon Europe (Version 2.0, 01 April 2022),

https://ec.europa.eu/info/fundingtenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

² The FAIR data principles (GO FAIR), <https://www.go-fair.org/fair-principles/>

During the course of the project, and at least at the moment of publication of project results, each research team will deposit and describe the relative underlying datasets. Trusted data repositories can attribute persistent unique identifiers (PIDs) to the deposited items. Readme files and the necessary documentation explaining all relevant details regarding data collection, processing methodologies, and quality assurance will be deposited alongside the datasets.

Specific keywords or variables used in the datasets will be, if possible, consistent with the vocabulary of the scientific field, and they will be uniquely defined in the metadata files provided alongside each dataset to ensure the interoperability of the data³. HYScale research data are organized in datasets, which are named collections of data units with the same focus and scope. This DMP suggests the following standard rules for dataset naming to improve data visibility, discoverability, citation, and permanent online tracking.

The recommended dataset title structure consists of the following:

PROJECT ACRONYM. WP number. WP title or description specifying WP aims. Task number. Task title or description specifying Task aims. Additional information specifying coverage and nature of data (if necessary). Version number (optional, in case of revisions to help identifying the updates especially in repositories that do not track versioning automatically).

The version number of the dataset will be added at the end of the title in case of data revisions to help identify the dataset updates, especially in repositories that do not track versioning automatically (see Annex I for dataset names, unique identifiers, and descriptions).

The DMP also recommends the following rules for file naming:

- for dataset file(s) [PROJECT ACRONYM]_WPnumber_Tnumber_coverage or other content specifications_date (YYYYMMDD)_vn.file extension
- for readme file(s)⁴ [PROJECT ACRONYM]_WPnumber_Tnumber_coverage or other content specifications_date (YYYYMMDD)_vn_README.file extension

WPnumber means “work package number”, Tnumber is “task number”, and vn is the “version number” (in case of data revisions or updates).

3.2 Making data openly accessible

As a guiding principle, HYScale seeks to ensure open access to research data via the repository as soon as possible and within the limits and deadlines set out in the DMP to allow dissemination, validation, and reuse of research results. In addition, data deposition in repositories will guarantee long time preservation and accessibility to datasets. Restrictions to access are applied only in the following cases:

- when collected data belongs to a third party which have denied permission to share them;
- on account of confidentiality and proprietary issues.
- protection of the personal data of subjects involved in the research.

³ 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.qo-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

⁴ A “README” file is a document containing relevant information about dataset authorship, terms of reuse and responsibilities, explaining dataset content and structure, collection procedures and analysis (such as file specifics, methodologies, codebooks of variables, data sources, and further necessary notes). (See Annex III to visualize the suggested README file template).

As a consequence, all possible and legitimate actions and strategies will be adopted to allow data sharing, including:

- converting the files to standard open formats;
- providing all relevant documentation and explanation for the data and the datasets;
- obtaining written consent to participate in the study and on the processing of personal data by each participant to the study;
- aggregating/anonymizing personal data.

For data that fall under some of the restrictions described above and for which it is not possible to take any action to make them shareable, the EU allows complete closure or restricted access to them. The main reason for limited access to data produced in the HYScale project will be the protection of intellectual property rights; an embargo will be applied for the time needed for the patent submission and/or to keep them as an industrial secret during the exploitation phase. No issue will derive from collecting personal data because none of them will be used in the research, because all the data will be anonymized before publication and/or sharing of the datasets. Each different dataset and software are deposited by the team responsible for the data collection and management in the repository of their choice. Research software code produced within the project will be deposited and made accessible through an open repository (such as GitHub (<https://github.com/>)) under a permissive license (such as MIT). Any release of the software will also be deposited on Zenodo (<https://zenodo.org/>), thus creating a persistent identifier ensuring findability. With regards to software, all design and development of software in this project will adhere to the ELIXIR Research Software Best Practices (taking into consideration the following aspects: (1) documentation; (2) testing; (3) interoperability to existing standards; (4) licensing; (5) version control; (6) attribution) and the FAIR software principles. For any additional outputs (such as documents, reports, presentations, posters, etc.); a dedicated project page will be created under Zenodo under a Creative Commons license for the deposit of outputs.

3.3 Making data interoperable

All datasets will be described using standard descriptive metadata (as an example, Dublin core metadata standard is supported by Zenodo) to ensure metadata interoperability for indexing and discoverability. Furthermore, all relevant documentation explaining codebooks, user manuals, data collection procedures and analysis will be made available along with the data to guarantee intelligibility, reproducibility, and the validation of the project findings. Partners will convert all shareable data from proprietary formats to allow data exchange and reuse among, e.g., researchers, institutions, organisations, and countries. Data will be made available in well-known and documented open formats, as much as possible compliant with available (open) software applications. The deposited documentation will include a full explanation and instructions.

3.4 Making data reusable

HYScale distributes the shareable data by adopting licenses that allow the reuse of the data and the datasets in their entirety by other scholars and stakeholders. Whenever the nature of the data allows it and at the appropriate timing, datasets will be made available under CC BY or CC0 licenses. In addition, an embargo period will be applied to some datasets to allow full exploitation of research results by the Partners. The quality of the data will be carefully assured by repeating the measurements several times and analysing errors and inconsistencies with

standard statistical methods. Besides their deposition in trusted repositories, which will grant their reusability by third parties after the end of the project, data will be given a full citation from official project publications and the project website.

5. Data security

Data shared among Partners will not contain sensitive data because they will be anonymized. Data generated by each partner will be initially stored in the computers, laptops, intranet directories, hard drives, and cloud storage systems of the research institution, accessible through institutional passwords modified periodically and protected by regularly updated antiviruses. None of the project data will be left inadvertently available by being left on desks or in unlocked rooms. All the research materials stored in computers are subject to back up regularly (according to each institution's regulations) to safeguard them from accidental losses. Cloud storage will be used for regular backup before data deposition in a repository. SharePoint will be set as the channel for data sharing amongst the partners. A team site has been implemented by the coordinator.

6. Private data collected via the HYScale events

The registration in HYScale events will have "Terms&Conditions" identified. To register and participate in such events, participants must agree to the set terms "Terms&Conditions", which may include the recording of the event. In any case, participants will be informed about the recording of the event prior to the start, leaving them the opportunity to be excluded from the recordings or photographs. When the events are conducted online/hybrid, they will also apply the used virtual meeting software's privacy policies. For such events, in general MS Teams and Zoom will be used.

Any personal information, including names, addresses, social media or email contacts, photos, voice recordings, and video, will be secured throughout the activity. The participants are allowed to choose how much private information to divulge to the other actors throughout the sessions and afterwards. Following the sessions, only the most necessary and authorised personal data will be disclosed to third parties. Otherwise and in general, personal data will be anonymised (where relevant), respecting GDPR⁵.

Each project data processor will save all participants' personal information offline or online on MS OneDrive and only use it to get in touch with them as a follow-up of the activity..

The information will subsequently be handled carefully by the data processor, who will take the required actions to turn the personal data into anonymous data, where relevant.

⁵ *General Data Protection Regulation (GDPR) – Official Legal Text.* (2022, September 27). General Data Protection Regulation (GDPR). <https://gdpr-info.eu/> [Accessed 13 April 2023].

The general data protection regulation. (2022, September 1). European Council. <https://www.consilium.europa.eu/en/policies/data-protection/data-protection-regulation/#gdpr> [Accessed 13 April 2023].

7. Ethics

The research activity in the HYScale project will not involve relevant ethical issues, nevertheless, it will be conducted by applying fundamental ethical principles and relevant national, EU, and international legislation, including the Charter of Fundamental Rights of the European Union, the European Convention on Human Rights, and the General Data Protection Regulation (GDPR – Regulation EU 2016/679). All lead investigators will be responsible for ensuring that ethical standards compatible with and equivalent to those of Horizon Europe will be applied, regardless of the country in which the research is carried out. As a general principle, all materials that could lead to an identification of the person are stored separately from actual data and handled by different research team members. No identifiable personal data will be considered for dissemination. Complete data anonymization will be performed before inclusion in the current datasets.

www.hyscale.eu

