D8.3 – DATA MANAGEMENT PLAN

WP8 – Task 8.3

Version 2

Dissemination level PU

03/04/2024

Authors

Hannah Johnson (TME)

Mina Lazarevic (LGI)

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

The project is supported by the Clean Hydrogen Partnership and its members.

Clean Hydrogen Partnership

Co-funded by the European Union

Document information

Grant Agreement	N° 101137889
Project Title	PH2OTOGEN
Project Acronym	PH2OTOGEN
Project Coordinator	Hannah Johnson, TME
Project Duration	M48
Related Work Package	WP8
Related Task(s)	T8.3
Lead Organisation	TME
Contributing Partner(s)	LGI
Due Date	31/03/2024
Submission Date	03/04/2024
Dissemination level	PU

History

Date	Version	Submitted by	Reviewed by	Comments
31/03/2024	1	Mina Lazarevic	Hannah Johnson	
03/04/2024	2	Mina Lazarevic	Hannah Johnson	

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

Table of contents

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

Summary

The main goal of the Data Management Plan is to describe how the research data will be handled during and after the end of the project, what kind of data will be collected, processed and/or generated, which methodology and standards will be applied, whether data will be shared/made open access and how data will be curated and preserved.

Any peer-reviewed research publications, abstracts and presentations for international conferences may also be shared through the website of the project at https://PH2OTOGEN.eu/

Keywords

Data Management Plan, datasets, open access, FAIR principles, Zenodo.

Abbreviations and acronyms

Acronym	Description
WP	Work Package
CA	Consortium Agreement
CC	Common Creative
D	Deliverable
DMP	Data Management Plan
DPO	Data Protection Officer
DoA	Description of Action
DOI	Digital Object Identifier
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
HE	Horizon Europe
IPR	Intellectual Property Rights
ISSN	International Standard Serial Number
M	Month (of the project)
NDA	Non- Disclosure Agreement
PII	Publisher Item Identifier
PMO	Project Management Office
PU	Public
ORDP	Open Research Data Pilot

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

SENS	Sensitive
URL	Uniform Resource Locator
V	Version
WP	Work Package
WPL	Work Package Leader
Gold Open Access	Open access publishing (gold open access) means that an article is immediately provided in open access mode on the publisher or journal's website. Some publishers charge Article Processing Chargers (APCs) to make articles open
Green Open Access	Self-archiving (green open access) means that a published article or the final peer-reviewed manuscript is archived (deposited) in an online repository before, alongside or after its publication. In some cases, the author can choose to delay access to the article (embargo period). HE rules state that embargo periods cannot exceed six months, except for publications in social science and humanities where the maximum embargo period is twelve months
Personal Data	Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected can lead to the identification of a particular person, also constitute personal data. Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the law. Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible
Zenodo	Zenodo is a catch-all research data repository that enables researchers, scientists, EU projects and institutions to share research results, make research results citable, and search and reuse open research results from other projects. Zenodo is harvested by the OpenAIRE portal and hosted by the CERN cloud infrastructure

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

Introduction

This deliverable provides an easy overview of data the PH2OTOGEN project is expected to generate, the types and formats of this data, and how this data is processed and stored to make them findable, accessible, interoperable and reusable, according to the principles of FAIR data management. The purpose of the DMP is to contribute to good data handling during the project's lifetime, and to describe how such data will be curated and preserved.

1 Purpose of the Data Management Plan

The purpose of this document is to lay out a plan for the management of data generated/collected in PH2OTOGEN. It covers the following:

- Identification of data to be collected/processed/generated
- Methodology and standards to be applied
- Data handling during and after the project
- Sharing, curating ad preserving data

At the time of this writing, PH2OTOGEN partners have identified 17 datasets to be included in the DMP at this stage. The majority of the datasets identified at this stage will be made openly accessible to the public through repositories such as Zenodo, and they will be preserved after the end of the project. The DMP is a living document – if necessary, it will be updated throughout the project's lifetime.

2 Data set properties

Following the guidelines of the EC (EC, 2016), this document contains the following properties for each of the identified data sets:

1. Name
2. Short description
3. Standards to be applied, metadata
4. Data sharing
5. Curation/archiving/preservation

A short description of each of these properties is provided below.

2.1.1 Name and reference code

In order to imbue the names of datasets with easily identifiable meaning that conveys important information, the following naming convention shall apply:

CountryCode.DataOwner.Openness.Title

CountryCode: this string identifies the country to which the data pertains/where the data was collected using the ISO 3166 Alpha-2 coding system.

DataOwner: this string identifies the project partner in PH2OTOGEN that is associated with the dataset (data collector/custodian) using the official abbreviated partner names.

Openness: this string determines whether a given dataset is intended to be shared with the public as Open Data. It may take the following values:

1. Open: can be accessed, used and shared by anyone without limitations, accessible on the internet in a machine-readable format, free of restrictions on use in its licensing)

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

2. Shared: available to use, but not under an open data license. Restrictions on its use or reproduction may apply (limited to a given group of people or organisations, may not be reproduced without authorisation, etc.)
3. Closed: can only be accessed by its subject, owner or holder

Title: a short and descriptive string to identify the contents of the data

Using these strings, the name of a dataset would look like this:

FR.LGI.Open.SolarenergycollectedBordeaux

A dataset with this name would describe a survey on the solar energy collected in Bordeaux area conducted in France and curated by LGI.

2.2 Data licensing

Without a license to set out the terms of use, data is not truly open. Data without a license may be publicly accessible, but users do not have the certainty that they can use and share the data, leaving them in a legal grey area. Data licensing standards are used to lay out the openness of data sets in concrete terms, and an open data license gives explicit permission to use the data both for commercial and non-commercial purpose. There are many types of licenses to choose from, and this document will not cover them in depth. The table below provides a summary of common data licenses that will be considered for use in the project (based on definitions from opendefinition.org):

Name	Domain	Attribution	Share-alike*	Notes
Creative Commons CCZero (CC0)	Content, data	N	N	All rights (including those of attribution) waived
Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	N	N	All rights (including those of attribution) waived
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, data	Y	N	Credit must be given, a link to the license must be provided, changes made must be indicated. If these terms are not followed, license may be revoked
Open Data Commons Open Database License (ODbL)	Data	Y	Y	Credit must be given, share-alike must be assured, data may be redistributed using DRM as long as a DRM-free version is also released

Table 1: Common Data Licenses

*Share-alike is the requirement that any materials created using the given dataset must be redistributed under the same license

2.3 Making data findable, including provisions for metadata

Digital Object Identifier (DOI)

All open data, publications and open source software produced in PH2OTOGEN will be identifiable and locatable by means of a persistent Uniform Resource Locator (URL). If

Clean Hydrogen Partnership

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

possible, open PH2OTOGEN results will be assigned a Digital Object Identifier (DOI) in order to make content easily and uniquely citable.

Open results that are deposited in the default Open Access repository (Zenodo, see below) will be assigned a DOI automatically and will benefit also from Zenodo's DOI versioning support.

Open results that are deposited in institutional repositories, repositories of scientific publishers or other data and research repositories will be at least definable by a persistent URI. If the institution is a DOI registrant that has an agreement with a DOI registration agency, a DOI will be assigned, too.

Whether scientific publications will be assigned a unique identifier like DOI, Publisher Item Identifier (PII), International Standard Serial Number (ISSN), etc. depends on the open access strategy (green or gold) chosen by the editors and thus also on the respective scientific publisher and the chosen research repository.

Zenodo

PH2OTOGEN open data will be collected in an open online research data repository: Zenodo. Its repository structure, facilities and management are in compliance with FAIR data principles. Zenodo is an OpenAIRE that allows researchers to deposit both publications and data, providing tools to linking them to these through persistent identifiers and data citations. Zenodo is set up to facilitate the finding, accessing, re-using and interoperating of data sets, which are the basic principles that ORD projects must comply with. Zenodo repository is provided by OpenAIRE and hosted by CERN. Zenodo is a catch-all repository that enables researchers, scientists, EU projects and institutions to:

- Share research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science;
- Display their research results and get credited by making the research results citable and integrating them into existing reporting lines to funding agencies like the European Commission;
- Easily access and reuse shared research results;
- Assign a DOI automatically to each result deposited in the repository,
- Integrate their research outputs with the OpenAIRE portal.

Search keywords

Zenodo allows to perform simple and advanced search queries on Zenodo using the keywords. Zenodo also provides a user guide with easy to understand examples. The Data Controllers at each pilot site will be responsible for uploading public datasets that they have generated and to assign specific keywords relevant to these datasets. Dataset specific keywords must be descriptive to the content of the dataset. In addition, the project has defined a set of general keywords that should apply to all public datasets, scientific publications and public deliverables. These are as follow:

Climate resilience, local communities, strategies, regions, inclusive, sustainable future.

Version numbers

Individual file names and datasets will contain version numbers that will be incremented at each revision (V*xyz*). For publications, versioning is in general not necessary.

Zenodo provides DOI versioning of all datasets uploaded to their communities, which allows us to edit and update the uploaded datasets after they have been published. This also allows us to cite specific versions of an upload and cite all versions of an upload.

Metadata

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

Metadata is data on the research data themselves. It enables other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers, can better determine whether the dataset is relevant and useful for their own research. Metadata will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data.

All metadata will be made available and is licensed under CC0, as soon as the different topic specific metadata schemas are established within the project. It is currently recognized within the project, that there are various gaps and conflicting metadata schemas and definition, which will require first a harmonization or possibly a distinction between the different fields.

All datasets, which will be shared openly on trusted repositories will be available for at least 10 years. This will be one of the criteria, which is in the checklist to assess any repository, before it is put on the list of trusted repositories, which should be used by project partners. Furthermore, it will be checked if the metadata will remain openly available, even after the datasets have been discarded after the data retention period has expired. This also depends on the arrangement of the repository to warrant functionality over a given period of time. Assessing this part will also be part of the checklist to assess repositories. This common repository checklist will be part of the next version of the DMP.

The following deposition metadata fields are mandatory for PH2OTOGEN:

- the terms “European Union (EU)” and “Horizon Europe”;
- the name of the action, acronym and grant number;
- the title and description of the deposition metadata,
- the upload type (publication, dataset, software, …),
- the publication date (ISO8601 format, YYYY-MM-DD), and length of embargo period if applicable;
- the creators/authors of the deposition,
- the persistent identifier (DOI),
- the used keywords.

This minimal metadata schema can be extended by arbitrary subjects upon request of the consortium.

2.4 Making data openly accessible

2.4.1 Data made openly available as the default

In order to maximise the impact of PH2OTOGEN research data, the results are shared within and beyond the consortium. Selected data and results will be shared with the scientific community and other stakeholders through publications in scientific journals and presentations at conferences, as well as through open access data repositories.

The PH2OTOGEN project datasets are first stored and organized in a database by the data owners (personal computer, or on the institutional secure server) and on the project database (project website). All data are made available for verification and re-use, unless the task leader can justify why data cannot be made openly accessible. To protect the copyright of the project knowledge, Creative Commons license will be used in some cases.

The PH2OTOGEN dataset deliverables are both public (data access policy unrestricted) and they will be accessible by:

- PH2OTOGEN project web site

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

- Partners database
- OpenAIRE
- Zenodo for ORDP data and datasets
- Open access journals

All data deposited on Zenodo are accessible without restriction for public.

2.4.2 How will the data be made accessible

PH2OTOGEN open results will be made accessible according to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon Europe.

Open data: All open results (data, software, scientific publications) of the project will be openly accessible at an appropriate Open Access repository (i.e. Zenodo) after publication in a scientific journal. Specifically, research data needed to validate the results in the scientific publications will be deposited in a data repository after the publication. Non-public research data will be archived at the repository using a restricted access option. Currently all data sets identified in the annex are closed due to partners' legitimate interests / potential for commercial exploitation i.e. to allow examination of the data for IP opport

Scientific publications: Providing open access to peer-reviewed scientific publications can be ensured either by publishing in green or gold open access journals with or without author processing fees. Any scientific publications from PH2OTOGEN and the related bibliographic metadata must be made available as open access and announced on the project website (https://gemini-initiative.com/) as well as in the OpenAIRE portal (https://www.OpenAIRE.eu/) and the R&I Participant Portal (https://ec.europa.eu/research/participants). To automate the process of reporting scientific publications and related research data in OpenAIRE, the publication should be deposited in an OpenAIRE-compliant repository, either by the authors of the publication (green open access) or by a scientific publisher (gold open access). While additional forms of disseminating open access papers, including academic social network sites such as ResearchGate (https://www.researchgate.net/) are possible, an electronic copy of the publication has to be deposited in suitable open access repository in the first place. According to the European Research Council's Guidelines on Open Access, "Venues such as Research Gate or Academia.edu that require users to register in order to access content do not count as repositories. The posting of publications on a personal, institutional or project specific webpage or the deposit in a publicly accessible Dropbox account is not sufficient to satisfy the requirements either."

If the chosen repository is not fully OpenAIRE complaint, the publications or data must be linked at https://www.openaire.eu/participate/claim with the respective funding agency (European Commission). Green open access journals or gold open access journals without author processing fees should be preferred for disseminating scientific publications of the PH2OTOGEN project. Nevertheless, the journal's visibility and prestige (translated in the Impact Factor) of the journal, together with the speed of publication, should be considered when choosing a journal for publication of a manuscript. According to the EC recommendation, authors of the publication are encouraged to retain their copyright and grant adequate licences to publishers.

- Green open access (self-archiving)
  Green open access or self-archiving means that the published article or the final peer-reviewed manuscript is archived by the researcher itself in an online repository, in most cases after its publication in the journal. The journal must grant the researcher the permission to self-archive the final peer-reviewed article, at the latest, 12 months after publication. For finding suitable green open access publishers, researchers are

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

encouraged to consult RoMEO (http://sherpa.ac.uk/romeo), a searchable database of publisher's policies regarding the self- archiving of journal articles on the web and in Open Access repositories.

– Gold open access (open access publishing)
Gold open access means that the publication is available by the scientific publisher as open access. Some journals require an author-processing fee for publishing open access. Author-publishing fees for gold open access journals can be reimbursed within the project period and budget. Some publishers allow the researcher to deposit a copy of the article in a repository, sometimes with an embargo period. For finding suitable gold open access publishers, researchers are encouraged to consult the Directory of Open Access Journals (https://doaj.org/), a service that indexes high quality, peer-reviewed open access academic journals that use an appropriate quality control system.

2.4.3 Methods and/or software needed to access the data

Regarding the mere access to open data deposited as data files in a data repository, there are no special methods or software tools needed. The files can be downloaded from the data repository using a standard web browser. The offline viewing, interpreting, processing and editing of data files downloaded from the data repository, it heavily depends on the type and format of the data.

2.4.4 Data sharing exceptions

1. Copyright and permissions for re-using third-party data sets

Processing and combining input data from many different sources may lead to unclear IPR situations regarding the generated output data, therefore such repurposed data (e.g. model output data) can only be made open if any of the underlying data (e.g. model input data) is open, too.

For this: potential users must contact the IPR (Intellectual Property Rights) team or the data owner in order to gain access. If necessary, appropriate IPR procedure (such as non-disclosure agreement - NDA) will be used.

2. Personal data treatment and confidentiality issues

Datasets referring to the quality and quantity of certain elements at risk, such as people and critical infrastructures, are not open by default as their publication may pose privacy, ethical or security risks.

3. Data-driven business model

Data that will be exploited commercially will not be made open.

3 Description of the data

The following detailed information sheet will be produced for every dataset to be produced/collected/curated in the project:

Name of the dataset	A name to identify the data, see 2.1.1 for details.
Description of the dataset	• A brief, easy to understand description of what the dataset contains and what it will be used for in the project • A list of institutions to whom the data set could be useful outside the project

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	• Whether the dataset has been/will be used for a scientific publication (if yes, brief details about the content and journal) • If the dataset is collected, a brief description of its origin and how it was collected will be provided • Openness of the dataset • Whether the dataset is anonymised or not.
Format/license	The format in which the data will be available (e.g. .xls, .csv, .txt) will be provided. The license to be used will also be provided.
Archiving/preservation	Efforts and means to keep the data available after the end of the project will be described here, including where/how the data will be preserved, the duration of preservation, the associated costs and the plans of the consortium to cover these costs.

Table 2: Dataset Example

4 Data Protection Officer

In accordance with applicable regulations, the host institution (TME) is not required to appoint a Data Protection Officer. A detailed data protection policy for the project is kept on file.

5 Ethical aspects

This Data Management Plan (DMP) was drafted taking into account the General Data Protection Rules (GDPR) for the collection, storage and re-use of the data, in line with the following general principles.

Personal data shall be:

1. processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
2. collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), however not be considered to be incompatible with the initial purposes ('purpose limitation');
3. adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
4. accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
5. kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
6. processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

6 Restrictions for re-use

Data generated through interviews and surveys will not be re-used directly due to privacy concerns. To allow re-use and avoid loss of research data, two different techniques could be used to disseminate its data while abiding by regulations on privacy.

6.1 Anonymization of data

“Anonymization" of data means processing it with the aim of irreversibly preventing the identification of the individual to whom it relates. Data can be considered anonymised when it does not allow identification of the individuals it is related to, and no individuals can be identified from the data by any further processing of that data or by processing it together with other information which is available or likely to be available.

There are different anonymization techniques. Here are the two most relevant:

- Generalisation: generalising data means removing its specificity. For example, in the case of a table containing household income levels, with 4 figures mentioned: €135,000, €60,367, €89,556, and €365,784. One way of generalising this numbers would be to write that the values are “more than €150,000, less than €80,000, between €90,000 and €120,000, and more than €300,000” respectively. Essentially it means taking exact figures, establishing a baseline category, and then obfuscating the data by assigning it to one of the categories in order to remove any sense of specificity from it.
- K-anonymity: a release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from the other individuals whose information also appear in the release. For instance, in a table composed of six attributes (Name, Age, Gender, State of Domicile, Religion and Disease), removing the name and the religion column while generalising the age is a way to effectively k-anonymise the data.

Other techniques, such as “masking” or “pseudonymisation”, which are aimed solely at removing certain identifiers, may also play a role in reducing the risk of identification. In many cases, these techniques work best when used together.

6.2 Pseudonymisation of data

"Pseudonymisation" of data means replacing any identifying characteristics of data with a pseudonym, or, in other words, a value which does not allow the data subject to be directly identified.

Although pseudonymisation has many uses, it should be distinguished from anonymization, as it only provides a limited protection for the identity of data subjects in many cases as it still allows identification using indirect means. Where a pseudonym is used, it is possible to identify the data subject by analysing the underlying or related data.

Task leaders will be responsible for the anonymization of data in PH2OTOGEN for all datasets where this is deemed necessary.

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

7 Personal data transfer and processing

In case personal data will be transferred from the EU to a non-EU country or international organisation, such transfers will be made in accordance with Chapter V of the General Data Protection Regulation 2016/679, and such transfers will comply with the laws of the country in which the data was collected.

In case of further processing of previously collected personal data, PH2OTOGEN will ensure that the beneficiary has legal grounds for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.

8 References

European Commission. ***Horizon Europe Programme****. Guidance for the classification of information in research projects*.

European Commission. ***Horizon Europe Programme****. Guidelines on FAIR Data Management in Horizon Europe*.

European Commission. ***Horizon Europe Programme****. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon Europe*.

European Union (2020). Intellectual property rights. https://europa.eu/youreurope/business/running-business/intellectual-property/rights/index_en.htm

European Parliament (2021). *Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32021R0695

European Parliament (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*. https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32016R0679

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

9 Annex I.: Datasets collected in the project

All detailed information sheets are below:

WP1

Name of the dataset	*BE.TME.Closed.HEP_Raw_Data*
Description of the dataset	Contains data on transition metal dichalcogenide HEPs: • UV-vis spectroscopy (absorption and transmittance) • Photoelectrochemical measurements performance and stability (linear sweep voltammetry, chronoamperometry, chronopotentiometry) • Photocatalytic measurements (gas evolution measurements) • Structural characterisations (X-ray diffractions, electron microscopy) • Chemical characterisations (FTIR, Raman, XPS) Raw data for submission of publications and patent filling
Format/license	The data will be available in the following format(s): .csv; .tiff The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Zenodo The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

Name of the dataset	*DE.FAU.Closed.HEP_Raw_Data*
Description of the dataset	Contains data on organic semiconductor HEPs: • Film production data form automated processing (pipetting, transport, film formation, timestamps) • Images of samples • UV-vis spectroscopy (absorption and transmittance) • Photoelectrochemical measurements performance and stability (linear sweep voltammetry, chronoamperometry, chronopotentiometry) • Optical characterisations (PL, TAS) Raw data for submission of publications and patent filling

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

Format/license	The data will be available in the following format(s): .json The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Project Data Server The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

Name of the dataset	*FR.CEA.Closed.PEC_Raw_Data*
Description of the dataset	Contains data on: • Photo-Electrochemical Performance • Photo-Electrochemical stability (chronopotentiometry) Raw data for submission of publications and patent filling
Format/license	The data will be available in the following format(s): The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☐ODbL ☒Other, please specify: Confidential at first. To be discussed at a later stage for accessibility for research purposes (ODbl license).
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: CEA internal servers The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A

Name of the dataset	*CH.EPFL. Closed.HEP_Raw_Data*
Description of the dataset	Contains data on organic semiconductor HEPs: • UV-vis spectroscopy (absorption and transmittance)

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	• Photoelectrochemical measurements performance and stability (linear sweep voltammetry, chronoamperometry, chronopotentiometry) • Photocatalytic measurements (gas evolution measurements) • Structural characterisations (X-ray diffractions, electron microscopy) • Chemical characterisations (FTIR, Raman, XPS) Raw data for submission of publications and patent filling
Format/license	The data will be available in the following format(s): .csv; .tiff The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Zenodo The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

WP2

Name of the dataset	*DE.HZB.OP_Closed_Raw_Data*
Description of the dataset	Contains data on: doped bismuth vanadate and/or tungstate based OPs: • UV-vis spectroscopy (absorption and transmittance) • Photo(electro)chemical measurements performance and stability (LSV, CP) under continuous and chopped illumination • Structural characterisations (X-ray diffractions, Scanning Electron Microscopy) • Chemical characterisations (XPS) • Gas and liquid product analysis (GC and HPLC) Raw data for submission of publications and patent filling
Format/license	The data will be available in the following format(s): .csv; .tiff The license used for this dataset: ☐CC0 ☐PDDL ☒ CC-BY-4.0 ☐ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories:

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

Name of the dataset	*FAU.OP_Closed_Raw_Data*
Description of the dataset	Contains data on organic semiconductor HEPs: • Film production data form automated processing (pipetting, transport, film formation, timestamps) • Images of samples • UV-vis spectroscopy (absorption and transmittance) • Photoelectrochemical measurements performance and stability (linear sweep voltammetry, chronoamperometry, chronopotentiometry) • Optical characterisations (PL, TAS) Raw data for submission of publications and patent filling
Format/license	The data will be available in the following format(s): .json The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Project Data Server The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

Name of the dataset	*CH.SLX.Closed.Paste_Formulations*
Description of the dataset	Contains data on: • OP paste/ink formulations • Printing/deposition methods Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): The license used for this dataset:

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	☐CC0 ☐PDDL ☐ CC-BY-4.0 ☐ODbL ☒Other, please specify: Confidential at first. To be discussed at a later stage for accessibility for research purposes (ODbl license).
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Project Data Server The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A

Name of the dataset	*CH.SLX.Closed.OP_Light_Soaking*
Description of the dataset	Contains data on: • Pictures of OP particle light soaking before and after test Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Project Data Server The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A

Name of the dataset	*NL.DIFFER.Closed_OP_Raw_Data*
Description of the dataset	Contains data on bismuth vanadate and/or tungstate based OPs: • UV-vis spectroscopy (absorption and transmittance) • Photo(electro)chemical measurements performance and stability (LSV, CP) under continuous and chopped illumination • Structural characterisations (X-ray diffractions, Scanning Electron Microscopy) • Chemical characterisations (XPS) • Gas and liquid product analysis (GC and HPLC) Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): .csv; .tiff

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	The license used for this dataset: ☐CC0 ☐PDDL ☒ CC-BY-4.0 ☐ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Zenodo The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

Name of the dataset	*BE.TME. Closed.OP_Raw_Data*
Description of the dataset	Contains data on transition metal dichalcogenide OPs: • UV-vis spectroscopy (absorption and transmittance) • Photoelectrochemical measurements performance and stability (linear sweep voltammetry, chronoamperometry, chronopotentiometry) • Photocatalytic measurements (gas evolution measurements) • Structural characterisations (X-ray diffractions, electron microscopy) • Chemical characterisations (FTIR, Raman, XPS) Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): .csv; .tiff The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Zenodo The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

Name of the dataset	*CH.EPFL. Closed.OP_Raw_Data*
Description of the dataset	Contains data on organic semiconductor OPs: • UV-vis spectroscopy (absorption and transmittance)

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	• Photoelectrochemical measurements performance and stability (linear sweep voltammetry, chronoamperometry, chronopotentiometry) • Photocatalytic measurements (gas evolution measurements) • Structural characterisations (X-ray diffractions, electron microscopy) • Chemical characterisations (FTIR, Raman, XPS) Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): .csv; .tiff The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Zenodo The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

WP3

Name of the dataset	*NL.DIFFER.Closed.Tandem_ Raw_Data*
Description of the dataset	Contains data on bismuth vanadate and/or tungstate based OPs: • Photo(electro)chemical measurements performance and stability (LSV, CP) under continuous and chopped illumination for tandem OP and HEP combinations • Structural characterisations (X-ray diffractions, Scanning Electron Microscopy) • Chemical characterisations (XPS) • Gas and liquid product analysis (GC and HPLC) Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): .csv; .tiff The license used for this dataset: ☐CC0 ☐PDDL ☒ CC-BY-4.0 ☐ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Zenodo The duration of the preservation will be: N/A

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

WP4

Name of the dataset	*CH.SLX.Closed.Large_Area_Supports*
Description of the dataset	Contains data on: SEM pictures of large area FTO coated supports Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Project Data Server The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A

Name of the dataset	*CH.EPFL. Closed.TPCS_Raw_Data*
Description of the dataset	Contains data on TPCS: • UV-vis spectroscopy (absorption and transmittance) • Photoelectrochemical measurements performance and stability (linear sweep voltammetry, chronoamperometry, chronopotentiometry) • Photocatalytic measurements (gas evolution measurements) • Structural characterisations (X-ray diffractions, electron microscopy) • Mechanical strength data (stress-strain curves) • Chemical characterisations (FTIR, Raman, XPS) Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): .csv; .tiff The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: Zenodo The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

WP5

Name of the dataset	*CH.LRESE.Closed.PEC_Modeling_Data*
Description of the dataset	Contains data from the computational model of the full reactor including: • Design of the reactor • Material characteristics used • Operational conditions (flow rates, pressures etc.) Raw data for submission of publication and patent filing
Format/license	The data will be available in the following format(s): The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: EPFL repository infoscience The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

Name of the dataset	*FR.CEA.Closed.Demonstration_Raw_Data*
Description of the dataset	Contains data on: testing of the demonstrator (electrochemical and analytical) Raw data for submission of publication and patent filling
Format/license	The data will be available in the following format(s): The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☐ODbL

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union

	☒Other, please specify: Confidential at first. To be discussed at a later stage for accessibility for research purposes (ODbl license).
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: CEA internal servers The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A

WP6

Name of the dataset	*BE.TME.Closed.LCA_TEA_raw data*
Description of the dataset	Contains data on: • Inventory used for LCA and TEA studies (material quantities, energy consumption, process steps, raw material costs) • Analysis data from LCA for Experts on the LCA and LCC of photoreactor components Raw data for submission of publication and patent filling
Format/license	The data will be available in the following format(s): .csv The license used for this dataset: ☐CC0 ☐PDDL ☐ CC-BY-4.0 ☒ODbL ☐Other, please specify:
Archiving/preservation	The data will be made available through the following platform(s) and/or repositories: The duration of the preservation will be: N/A Foreseen costs of the preservation: N/A Means to cover preservation costs: N/A If the dataset contains personal data, will it be anonymised? N/A

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

The project is supported by the Clean Hydrogen Partnership and its members.

Co-funded by the European Union