



Demonstration of Sustainable Hydropower Refurbishment

D10.3 Data Management Plan



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101147310

| | |
|---------------------|-----------------------------|
| Deliverable No. | 10.3 |
| Deliverable Title | Data Management Plan |
| Work Package Title | Coordination and management |
| Dissemination level | Public (PU) |
| Due date | 2024-10-31 |
| Version | 1.0 |
| Status | Final |
| Submission date | 2024-10-31 |

| Contributors | Name | Partner | Date |
|------------------------------|---|--|------|
| Deliverable Leader | Atle Harby | SINTEF | |
| Work Package Leader | Arnt Ove Eggen | SINTEF | |
| Contribution Author(s) | Claire Ségoufin Serdar Kadam Alexander Jung Henning Lysaker Jean-Christophe Marongiu Michael Belsnes Frode Agedal | GE ANDRITZ VOITH AKER AND_CH SINTEF INTOTO | |
| Reviewer | Elena Vagnoni Cécile Münch-Alligné | EPFL HES-SO | |
| Final review and approval | Atle Harby | SINTEF | |

DISCLAIMER / ACKNOWLEDGMENT

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 the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101147310.

This report is licensed under a Creative Commons Licence.

Attribution 4.0 International License



For more information about the Creative Commons Licence, including the full legal text, please visit: <https://creativecommons.org/use-remix/cc-licenses/>.

Table of Content

| | | |
|-------|---|----|
| 1 | Data Summary | 6 |
| 1.1 | Purpose of the data collection and generation | 6 |
| 1.2 | Data types, formats and size | 6 |
| 1.2.1 | Types of data | 6 |
| 1.2.2 | Data formats | 6 |
| 1.2.3 | Size of data | 7 |
| 1.3 | Origin of data and provenance | 7 |
| 1.4 | Data storing in active project – project platform | 7 |
| 1.5 | Data Repository | 7 |
| 2 | FAIR data | 9 |
| 2.1 | Making data findable, including provisions for metadata | 9 |
| 2.1.1 | ReHydro community in Zenodo | 9 |
| 2.1.2 | Metadata in Zenodo | 9 |
| 2.1.3 | Versioning and Digital Object Identifiers (DOI) | 10 |
| 2.1.4 | Approach to search keywords | 10 |
| 2.1.5 | Naming conventions | 11 |
| 2.2 | Making data accessible | 11 |
| 2.3 | Making data interoperable | 11 |
| 2.4 | Increase data re-use | 12 |
| 2.4.1 | Recommended Creative Commons (CC) licences | 12 |
| 2.4.2 | Availability and longevity of ReHydro research data sets | 12 |
| 3 | Other research outputs | 12 |
| 4 | Allocation of resources | 13 |
| 5 | Data security | 13 |
| 5.1 | Active Project - Data security as specified for SINTEF SharePoint | 13 |
| 5.2 | Repository - Data security as specified for Zenodo | 14 |
| 6 | Ethics | 14 |
| 7 | References | 15 |

Executive Summary

The main objective of ReHydro is to demonstrate how European hydropower can be refurbished and modernized to be fit for a leading role in the future power system respecting sustainability requirements and societal needs in a climate change context. A suite of monitoring and digital tools (performance, cavitation, machine health) implemented at demonstration sites will improve hydropower efficiency. Innovative concepts like retrofitting with pumped hydro and hybridization will make hydropower fit for future markets.

This document is the Data Management Plan (DMP) for ReHydro. The main purpose of the DMP is to describe procedures related to data management in the project. This plan is related to ReHydro's Grant Agreement [1] and Consortium Agreement [2].

This documents' structure follows the Data Management Plan (DMP) template¹ from the European Commission and falls into dissemination level public, but will serve for internal use only, meaning only for members of the consortium and the granting authority.

The purpose of the DMP is to contribute to good data handling by describing what research data the project expects to use, data handling principles, and an assessment of what data can be shared with the public or why data cannot be open. Furthermore, it gives instructions on naming conventions and metadata structure.

Ethical aspects related to data collection, generation and sharing have been considered. All datasets will be uploaded, stored and handled in accordance with national and European rules on data protection and privacy. Metadata will be added to all datasets, and instructions on how to upload, store, publish and preserve research data is provided in this document.

ReHydro will use Zenodo, a trusted and OpenAire compliant data repository to comply with the FAIR data principles. All data sets with dissemination level "Public" will be uploaded to the ReHydro communities hosted on Zenodo (<https://zenodo.org/communities/rehydro/>).

Each dataset will be given a persistent identifier (Digital Object Identifier, DOI), supplied with relevant metadata, and linked to the project acronym, full project name and grant agreement number. Publications and underlying research data will be linked, and a Creative Commons license will regulate reuse of the ReHydro research data. Data security arrangements are defined.

Due to confidentiality reason related to design of electro-mechanical parts, patents, power-system sensitive information or other reasons, some data cannot be public or shared with the whole consortium. Such data must be managed, stored and shared according to the requirements of the owner of the data. For sharing such data, the owner of the data will make necessary agreements or contracts to maintain necessary sensitivity, security and access rights.

The Data Management Plan will be reviewed and updated regularly throughout the project's lifetime. The final version will include instructions for how to access and reuse open research data from ReHydro.

¹ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx

1 Data Summary

1.1 Purpose of the data collection and generation

The overall motivation for data collection in ReHydro is to be of value for use by all partners in ReHydro, as well as for other projects and research endeavours by:

- Facilitating evaluations and learning
- Evaluating measurements, results and solutions
- Validating measurements, results and solutions
- Gathering feedback to improve results and solutions
- Increase user acceptance
- Encourage behavioural change

More specifically data will be re-used to facilitate validation of the experimental measurements, simulations of the system and infrastructure, validation of the models, cross-checking information and to ensure the quality of the outputs.

Only data that is needed to perform project activities will be collected, and as far as possible, participants will not be asked to provide personal data unless this is necessary.

1.2 Data types, formats and size

1.2.1 Types of data

A preliminary overview of data types/outputs re-used and shared is given under by data type:

- **Experimental:** Validated data and measurements from experiments.
- **Images:** Drawings of parts as required as input for simulations by partners, detailed design plans of system layout and infrastructure, documentation of system.
- **Text:** Data generated within ReHydro, detailed in text format in deliverables and other reports.
- **Numerical:** Validated data from numerical simulations.
- **Publications:** Scientific publications will be published based on results from the research work in ReHydro, with a link to already existing publications from other projects.

Data will be organised in datasets according to type and content of the data.

1.2.2 Data formats

ReHydro aims at using widely accepted formats for data generation, such as:

- Documents, reports and publications (txt, docx, pdf)
- Spreadsheets (xlsx)
- Databases (csv)
- Audio files (mp3, wav, wma, ra)
- Pictures (jpg, png)
- Video (avi, flv, mov, mp4, wmv)
- Calculation data (binary)

- CAD geometries (x_t, stp, igs, dwg)
- Compressed archive (zip, rar)

1.2.3 Size of data

Due to the large amount of data that will be collected during field work, experiments, modelling, analysis and testing, the expected overall size of data is above 100 TB.

Other data platforms would be considered depending on the size of the data obtained during the project.

1.3 Origin of data and provenance

Each partner background is defined in the Consortium Agreement [2].

1.4 Data storing in active project – project platform

All non-confidential data that are relevant to other partners, will be stored in a SINTEF SharePoint project site. This will be the projects online working and collaboration area during the 48 months the project is active and open to all partners until 12 months after the end of the project.

Confidential data, raw data that are not made available on standardised formats or quality assured as well as data that are not relevant to other partners, will be stored by the partner owning the data according to the normal procedures for data storage within the partner organisation.

All partners will be responsible for uploading quality-assured datasets they have collected/generated during the project to a dedicated research data folder in the SharePoint site using standardised formats. All datasets will use standard SharePoint version control and access control to enable limited access to certain types of data. Each dataset will be catalogued in an Excel file “ReHydro Datasets Overview.xlsx” providing meta-data and an overview of all datasets in the project.

These metadata will be provided for each data set:

- Dataset name
- Date/time
- Data owner
- Version
- Description
- WP number
- Dissemination level

1.5 Data Repository

ReHydro will use Zenodo, a trusted and OpenAire compliant repository to comply with the FAIR data principles.

All scientific publications, including public deliverables and public parts of underlying datasets will be uploaded to the ReHydro repository (<https://zenodo.org/communities/rehydro/>). In

addition, the project will upload other datasets (not directly linked to publications and deliverables) with dissemination level "Public" and make them openly accessible via the repository.

2 FAIR data

ReHydro will manage data in accordance with the principles of FAIR data management² (Findable, Accessible, Interoperable and Re-usable data). The project aims to maximise access to, and re-use of research data generated by the project.

At the same time, there are datasets, or parts of datasets, generated in this project that cannot be shared:

- in order to protect the privacy of voluntary participants in the demo-sites
- to allow for protection of results prior to commercial exploitation
- to protect business sensitive information
- to protect confidential/classified information

2.1 Making data findable, including provisions for metadata

2.1.1 ReHydro community in Zenodo

ReHydro will use the Zenodo repository as the main tool to make our research data findable in accordance with the FAIR data principles. A ReHydro community has been established in the repository (<https://zenodo.org/communities/rehydro/>), and the project will upload all our public datasets and deliverables as well as scientific publications to this community. This will ensure harvesting of project results by OpenAire and provide maximum findability. All uploads will be enriched with standard Zenodo metadata, including Grant Agreement number, project name and acronym. The repository provides version control and assigns DOIs to all uploaded elements.

2.1.2 Metadata in Zenodo

Metadata associated with each published data set in Zenodo will by default be as follows³:

- Resource type
- Digital Object Identifier (DOI)
- Title
- Publication date
- Creators (Name and ORCID)
- Description
- Licenses and rights
- Contributors

In addition, the project name and Grant Agreement number would be added.

² https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

³ <https://help.zenodo.org/docs/deposit/describe-records/>

2.1.3 Versioning and Digital Object Identifiers (DOI)

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.

2.1.4 Approach to search keywords

Each partner who collects and generates public datasets will be responsible for uploading these to the repository and to assign specific search keywords relevant for 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 apply to all public datasets, scientific publications and public deliverables. These are as follows:

- ReHydro
- Hydropower
- Engineering
- Turbomachinery
- Power electronics
- Digital twin
- Numerical simulations
- Numerical analysis
- Electricity market
- Capacity market
- Day-ahead market
- Ancillary services
- Hybridisation
- Hydrology
- Hydraulics
- Biology
- Ecology
- Geomorphology
- Topographical data
- Fish migration
- Fish habitat
- Sediment transport
- Sediment erosion
- Water temperature
- Water depth
- Water velocity
- Socio-economic
- Social acceptance
- Remote sensing
- Environmental DNA (eDNA)

Other appropriate keywords may be added later.

2.1.5 Naming conventions

Data will be named using the following naming conventions:

Descriptive text_ReHydro_DeliverableNumber_UniqueDataNumber

Descriptive text_ReHydro_PublicationNumber_UniqueDataNumber

Explanation of the naming convention:

- "Descriptive text" refers to a short description of the content of the dataset (see example).
- "ReHydro" refers to the project's acronym.
- "DeliverableNumber" refers to the deliverable number as described in the DoA.
- "PublicationNumber" refers to the number the publication has in the project internal directory of all publications from the project.
- "UniqueDataNumber" is the number automatically generated by the research metadata list in SharePoint.

2.2 Making data accessible

The EU's Open Science Policy aims to make research data generated by Horizon Europe projects accessible with as few restrictions as possible, but also accept protection of personal or sensitive data due to privacy concerns and/or commercial or security reasons.

All public datasets, scientific publications and deliverables will be uploaded to Zenodo and made openly available, free of charge. Publications and underlying data sets will be linked through use of persistent identifiers (DOI versioning). Data sets with dissemination level "confidential" (non-anonymous datasets) will not be shared due to privacy/security/ethical concerns. Potentially, some datasets might be restricted due to protection for commercial exploitation. If such cases arise during the project, this will be informed in the final version of the DMP.

Metadata for individual data records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name. Metadata is also retrievable through the public REST API ⁴. The data will be available through www.zenodo.org, and hence accessible using any web browsing application. No specific software or hardware is needed to access and reuse the data.

2.3 Making data interoperable

Zenodo uses JSON Schema as internal representation of metadata and offers export to other formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and export to Mendeley ⁴. The data record metadata will utilise the vocabularies applied by Zenodo. For certain terms, these refer to open, external vocabularies, e.g.: license (Open Definition), funders (FundRef) and grants (OpenAIRE). Reference to any external metadata is done with a resolvable URL.

⁴ <https://about.zenodo.org/principles/>

2.4 Increase data re-use

ReHydro will enable third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) all public data sets, and regulate this by using Creative Commons Licences.

2.4.1 Recommended Creative Commons (CC) licences

ReHydro will use Creative Commons licences (CC), which are tools to grant copyright permissions to creative work. As a default, the CC BY license will be applied for public ReHydro research data. This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you.

This does not preclude the use of less restrictive licenses as CC 0 or more restrictive licenses as CC BY-ND, which does not allow derivations.

Application of licences will be assessed on a case-by-case basis in close collaboration with the coordinator and partners concerned.

2.4.2 Availability and longevity of ReHydro research data sets

For data published in scientific journals, the underlying data will be made available no later than by journal publication. The data will be linked to the publication. Data associated with public deliverables will be shared once the deliverable has been approved and accepted by the EC. For other public datasets not directly linked to a scientific publication or deliverable, such datasets will be made available upon assessment by the responsible partner that it is ready for publishing, and in the final month of the project at the latest.

Open data can be re-used in accordance with the Creative Commons licences. Data classified as confidential will not be re-usable due to privacy/security concerns.

The public data will remain re-usable via Zenodo for at least 20 years. This is currently the lifetime stated by the host laboratory CERN. If Zenodo must close their operations, they have provided a guarantee that they will migrate all content (including metadata) to other suitable repositories.

3 Other research outputs

ReHydro plan to produce digital and physical project results that will be openly shared and available. Such products will normally be included in project deliverables. If this is not the case, ReHydro website as well as social media posts will be used to inform about the results outside ReHydro partners, and any data, software, protocols or guidelines will be uploaded to Zenodo unless it is not public.

4 Allocation of resources

ReHydro uses standard tools and a free of charge research data repository. The costs of data management activities are limited to project management costs and will be covered by allocated resources in the project budget.

Long-term preservation of the public data is ensured through Zenodo. Other resources needed to support re-use of data after the project ends will be solved on a case-by-case basis.

Work package leader for WP10 and WP11 Arnt Ove Eggen (arnt.o.eggen@sintef.no) will be responsible for the data management and quality assurance.

5 Data security

In this chapter, the security features of the research data infrastructure used to store and handle data in ReHydro are described.

5.1 Active Project - Data security as specified for SINTEF SharePoint

SINTEF SharePoint is the online collaboration platform used by ReHydro. A dedicated project site has been established on this platform, accessible only by the partner representatives in the consortium. Furthermore, a dedicated folder for research datasets is set up, allowing for stricter access control than the main project site.

The ReHydro Sharepoint site has the following security settings:

- Access level: Restricted to persons (project members only). Further access restrictions on specific folders could if necessary be enabled.
- Encryption with SSL/TLS protects data transfer between partners and the SINTEF SharePoint site.
- Threat management, security monitoring, and file-/data integrity prevents and/or registers possible manipulation of data.

Documents and elements in the SINTEF SharePoint sites are stored in Microsoft's cloud solutions, based in Ireland and the Netherlands. There will be no use of data centres in the US or outside EU/EEA and associated countries.

Nightly back-ups are handled by SINTEF's IT operations contractor. As a baseline, all project data will be stored for 10 years according to SINTEF's ICT policy, unless otherwise agreed in contracts and data processing agreements.

5.2 Repository - Data security as specified for Zenodo

The following list describes the security settings for Zenodo⁵:

- **Versions:** Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- **Replicas:** All data files are stored in the CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- **Retention period:** Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
- **Functional preservation:** Zenodo makes no promises of usability and understandability of deposited objects over time.
- **File preservation:** Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- **Fixity and authenticity:** All data files are stored along with an MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
- **Succession plans:** In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

6 Ethics

The proposed work in ReHydro will fully comply with the regulations set out in Regulation (EU) 2016/679, the General Data Protection Regulation (GDPR) [3]. In addition, ReHydro comply with the principles of the European Charter for Researchers, the European Code of Conduct for Research Integrity, including ethical standards and guidelines, regardless country in which research is carried out.

Nothing in this project shall be deemed to require a party to breach any mandatory statutory law under which the party is operating. This includes any national or European regulations, rules and norms regarding ethics in conducting research.

SINTEF follows the Vancouver Recommendations for publication of scientific work.

An e-mail address is by definition personal information and covered by GDPR. The e-mail addresses of project participants are stored in the ReHydro SharePoint Site. The e-mail address is a prerequisite to access the projects working area. By accepting the SharePoint invitation, the participants consent to use and store their e-mail address for the purpose of online collaboration in the project. Only the project participants invited will have access to the site.

⁵ <https://about.zenodo.org/policies/>

The email-addresses will be deleted when access to the project area is no longer needed (12 months after project closure).

SINTEF has signed GDPR data processing agreements with both Microsoft and the IT operations contractor handling our SharePoint platform.

7 References

- [1] ReHydro Grant Agreement, grant agreement number 101147310.
- [2] ReHydro Consortium Agreement.
- [3] 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 (General Data Protection Regulation). Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>.
- [4] D1.4 Data management plan (DMP) from the project «HyPowerGT - Demonstrating a hydrogen-powered gas-turbine engine fuelled with up to 100% H₂».

