Pre-normative REsearch for Safe use of Liquid Hydrogen (PRESLHY)

Fuel Cells and Hydrogen Joint Undertaking (FCH 2 JU)

Grant Agreement Number 779613

Data Management Plan Version 1.4

Deliverable Number:	1.3
Title:	Data Management Plan Version 1.4
Authors:	T. Jordan
Submitted Date:	2 October 2018
Due Date:	30 June 2018
Report Classification:	Public

History		
Nr.	Date	Changes/Author
1.0 - Draft	2.10.2018	Thomas Jordan
1.1 - Draft	9.10.2018	Major editorial/language correction / Review by Simon Coldrick
1.3 - Draft	8.10.2018	Introduction of data management group, removing movies as potential data stream..
1.4	11.10.2018	Synthesis of review (Version 1.1 - Draft) and 1.3 -Draft

Approvals			
Version	Name	Organisation	Date
1.1 - Draft	Simon Coldrick	HSL	16 October 2018

Key words

FAIR data management, pre-normative research, experimental data, accessibility, re-use, long-term data storage, research data repository, liquid hydrogen, accidental behaviour, release, ignition, combustion

Acknowledgements, Preface and Disclaimer

This project has received funding from the Fuel Cells and Hydrogen 2 Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 779613

The data management in the PRESLHY project follows the principle of data management, which shall make data Findable, Accessible, Interoperable and Re-usable (FAIR). The plan for FAIR data management as described in this document is based on the corresponding template for open research data management plan (DMP) of the European Research Council (ERC).

Publishable Short Summary

The data management in the PRESLHY project follows the principle of FAIR data management. The plan itself is based on the associated template for open research data management plan (DMP) of the European Research Council (ERC)[1].

The plan describes the objectives of the project and the purpose of the generated data, which is mainly for validation of models describing the accidental behaviour of liquefied hydrogen. The data generated from release, ignition and combustion experiments will be edited and stored in a public repository together with its metadata for at least 10, if affordable even more than 25 years.

Re-use of data will be stimulated by publishing benchmarks and scientific papers referring to the corresponding data.

[1], https://erc.europa.eu/content/erc-data-management-plan-template

Table of Contents

1 Summary

The PRESLHY project will undertake pre-normative research for the most relevant and poorly understood phenomena related to high risk scenarios from transport, use and storage of liquid hydrogen (LH2). The project involves collaborative research work, consisting of experimental and theoretical investigations in the following key areas: flashing multiphase, multicomponent releases with intense phase changes, cryogenic, plumes and jets, ignition phenomena, such as electrostatic charge build-up in the presence of condensed or frozen phase, potential for flame acceleration and deflagration-detonation-transition in multiphase, multicomponent mixtures.

Standards and specific recommendations for transport, use and storage of liquid hydrogen are increasingly in demand in various industries, such as energy, automotive, and aerospace. To provide a solid risk-informed basis for these standards a good understanding and modelling capability of the above phenomena is required.

The primary objective of the experimental work is to provide fundamental data for development and validation of models, ranging from engineering type correlations to more sophisticated models for integration in space and time resolving multi-physics/CFD software. Only the data generated in the extensive and expensive experimental program is considered for the data management in the PRESLHY project.

The experimental program itself is split into three main parts treated in three different work packages: the first part deals with accidental release and mixing of the cryogenic hydrogen (WP3), the second deals with ignition phenomena (WP4), the third part with combustion behaviour (WP5). All experiments will be executed in the experimental facilities of the project partners KIT, HSL and INERIS.

Typical measurements undertaken in these experiments are hydrogen concentrations, pressure, temperature, flow velocities, mass and mass fluxes, displacements, acoustic levels, electric signals such as voltages and currents. These may be recorded at several locations and at high frequency. Optical measurements represent a key technology for deriving essential information from the experiments. Conventional photography, laser techniques and high speed movies capture important information about density variation and chemical reactions at critical instances and/or over time. Therefore the data will have the typical standardised formats used for storing floating point data and pictures and potentially movies (.CSV, .TXT, .PDF, .RAW, .JPEG, .CIF, .TIFF, .GPL, .AVI etc.). Actual formats and qualities will be determined during the final specification and installation of the respective experiments.

In particular optical data acquisition and management is demanding regarding formats and storage size. According to experience at KIT a typical size for a dataset derived from a single experiment is about 10 GB. With the foreseen number of tests in WP3, WP4 and WP5 the total size amounts to about 2 TB.

The large amount of data will be reduced by filtering, further aggregation and mainly by selecting the most relevant data for long term data storage. The selection process will account for quality assurance in the experimental procedures and for the priorities identified in the project.

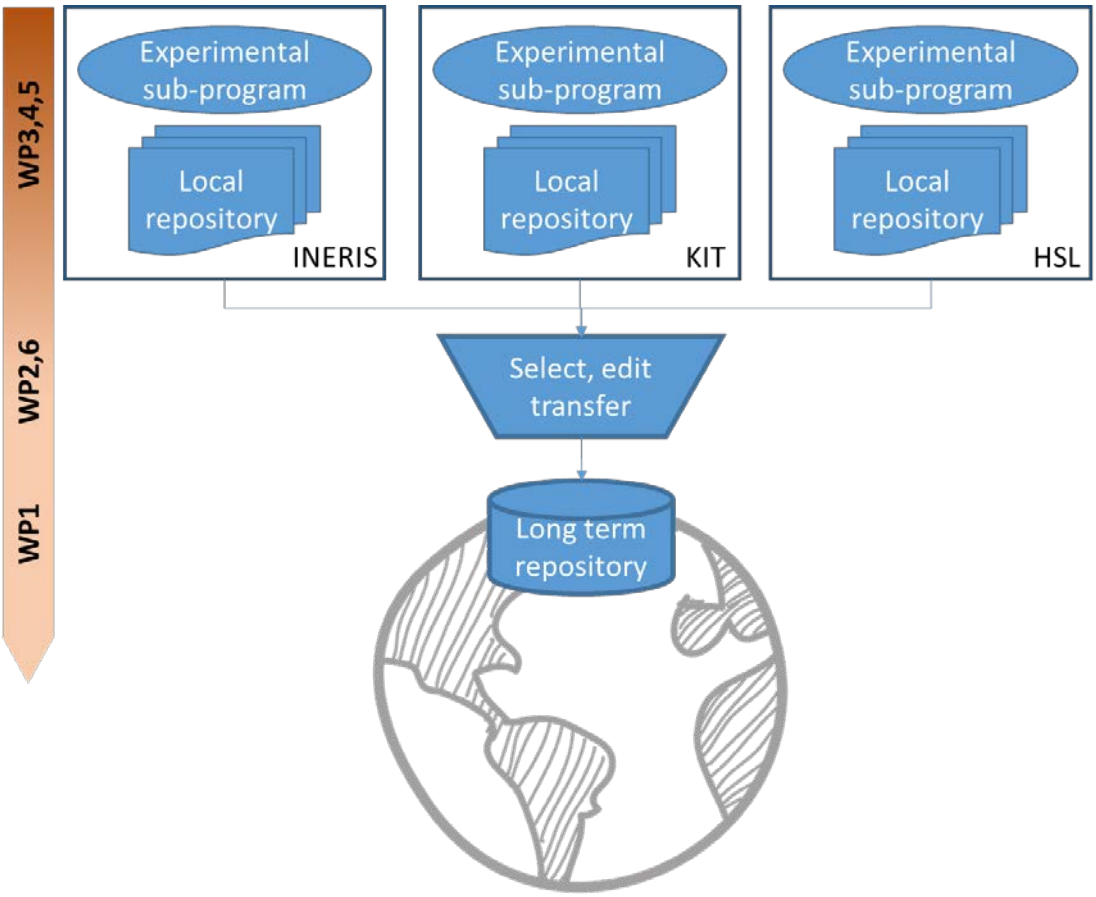

Fig. 1: Data management in PRESLHY project

The preserved data will be those deemed to be most useful for the research community for the further development and, in particular, the validation of models for the complex behaviour of LH2 in accidents. The data generated by the project therefore documents the advanced state-of-the-art. For industry involved in the development and marketing of innovative hydrogen energy solutions these data will help to refine their hazard and risk analysis procedures. Finally, for the general public stakeholders these data will have a transparent and scientific basis, helping to improve and to demonstrate the intrinsic safety performance of LH2 and the associated technologies.

2 Making Data Findable

A large amount of data will be generated from the experiments and simulations in PRESLHY. As the simulation results are deemed to be easily reconstructed the actual focus on data management is on the experimental data. The main participants in the project conducting experiments are: French National Institute for Industrial Environment and Risks (INERIS, France), Karlsruhe Institute of Technology (KIT, Germany) – largely represented by Pro-Science GmbH (PS, Germany) and Health & Safety Laboratory (HSL, UK). Those partners have their own established procedures for storing data locally.

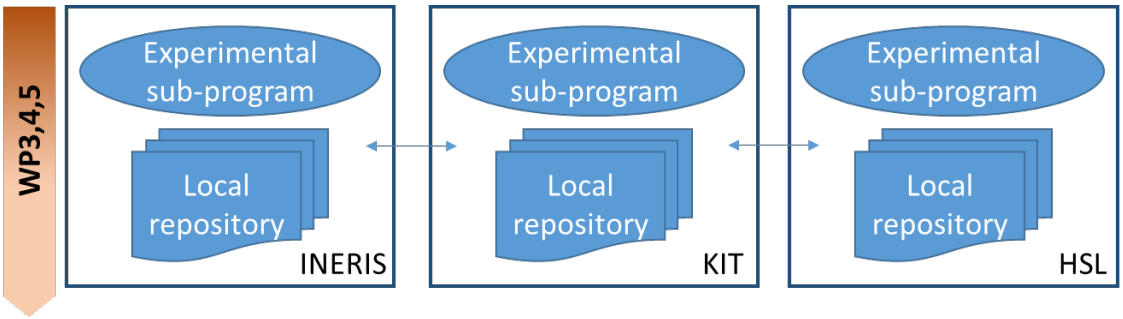

Fig. 2: Data exchange on project level

However, for the simplified further processing and making data easily findable the concerned partners agree to adhere to the file naming convention with a fixed length of 24 characters plus standard file extension as described in the figure below.

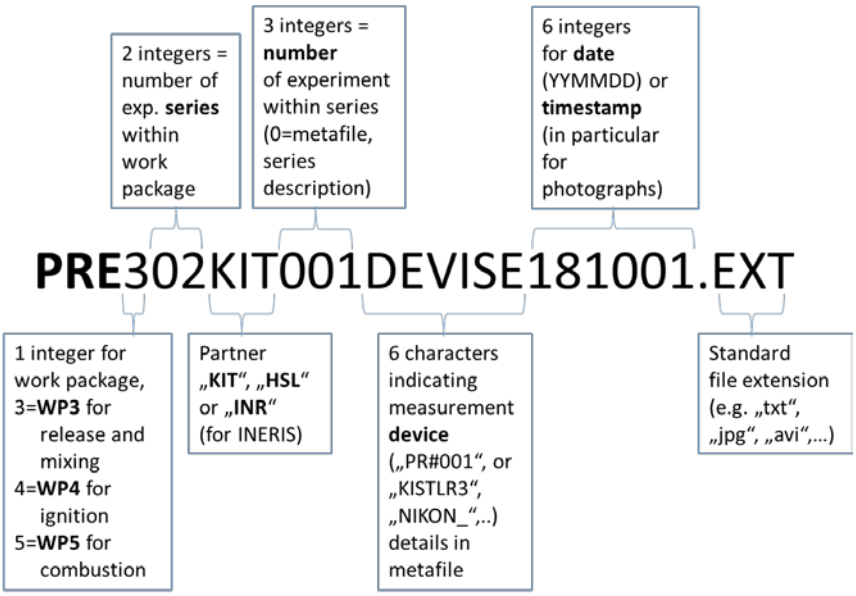

Fig. 3: File name convention, file or data identifier respectively

The first nine characters identify the test series number of a certain work package (WP3, WP4 or WP5) and the partner executing the actual experimental series. The work package and the test series number correlate with a similar naming scheme used in the PRESLHY description of work. The experiment referred to in Fig.2 would be E3.2 of the description of work.

Any set of result files containing measurement data, and optical data respectively, with identical first nine characters shall be accompanied by one metafile containing information about the whole experiment series and explaining the differences defining the different actual experiment, indicated by a number.

A typical metafile should give the metadata in descriptive form and should include essential elements such as the title and objective of the experimental series in short, names of the experimentalists, keywords etc. It shall describe the experimental set-up, where drawings or photographs are very instructive and shall provide details of the actual

test conditions (ambient pressure, temperature, relative humidity, etc.). Further details of a prototypical test series of combustion type (WP5) could be, but are not limited to:

- Orifice size, location and orientation
- Orifice shape
- Pressure at orifice
- Material of the ground
- Ignition source, location and timing

The metafile shall provide the experiment matrix, usually represented in a list of each associated experiment with respective number, deviation from the general description provided, and actual date of measurement. Furthermore, it shall list all measurement devices used with actual placement in the experimental set-up and respective measurement precision.

As the actual device indicated in the filename could generate more than one signal or for more compact storage several signal of different sensors might be packed in one result file, the device may be denoted as “MIXED_” and the required metadata has to be provided in the actual result file (in the header line for instance)

The date contained in the last 6 characters represents the date of measurement, the date of last data editing respectively. This date therefore implicitly represents the version number of the data. Any modification of the data should be explained in the metafile referring to the date of modification.

At project level, data might be exchanged at this early stage without further checks or evaluations.

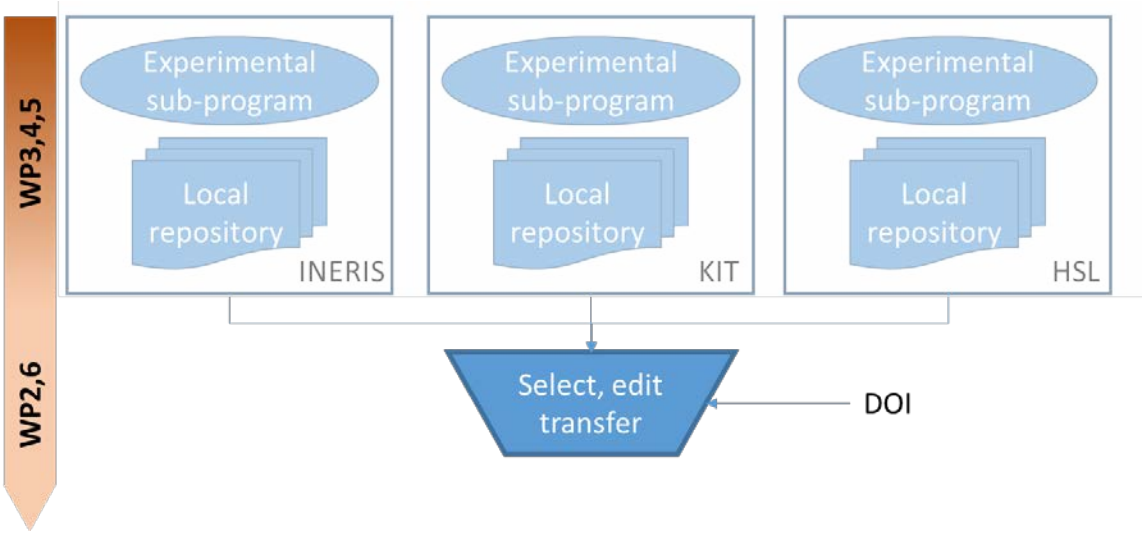

Fig. 4: Common data selection and processing

At the regular project meetings, the partners involved in strategy and dissemination will review new data and select most valuable results. In this stage, data and metadata will be checked, processed, reformatted and associated with a Digital Object Identifier (DOI) such that everything is in accordance with this Data Management Plan and with DOI standards.

Finally, these processed data will be uploaded on either the RADAR Research Data Repository[2] or the KITopen Research Data Repository[3].

3 Making Data Openly Accessible

All selected data will be fully accessible via the RADAR Research Data Repository or the KITopen Research Data Repository. Access to these repositories is free and supported by any web browser. The address of the repository will be advertised in all related publications and dissemination events.

As only standard and in most cases readable (ASCII) file formats will be used no specific software will be required. The only constraint might be imposed by the video data, in case requiring specific codecs and appropriate video player software. Therefore, for saving storage space selecting specific frames and storing them as picture data shall be the preferred strategy.

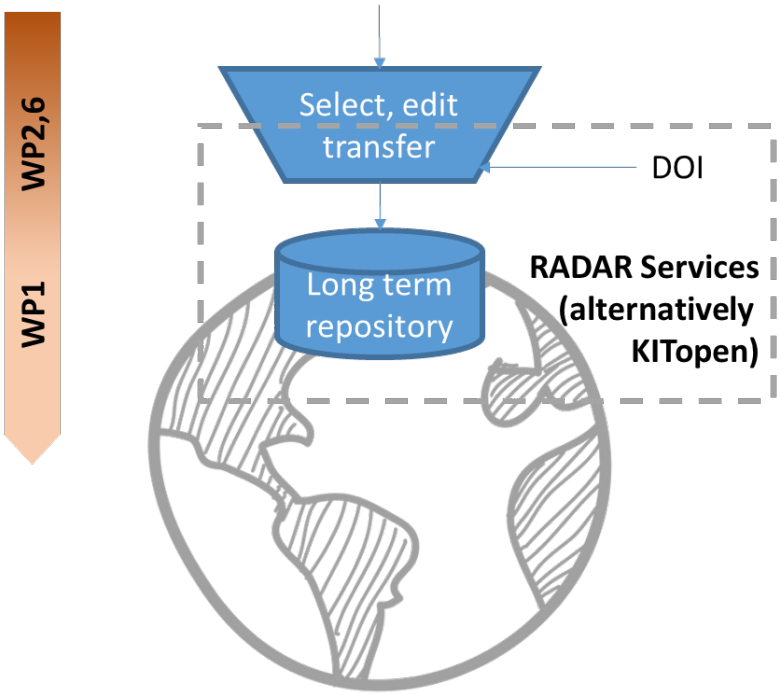

Fig. 5: Publishing data

[2] https://www.radar-projekt.org/display/RE/Home

[3] https://www.bibliothek.kit.edu/cms/english/kitopen-how-publish.php#Anker5

4 Making Data Interoperable

For the data file formats, only standard file formats will be used.

The PRESLHY members will use SI base units for any measured quantity. This implies that any units may be easily composed by multiplying powers of metre (m) for length, kilogram (kg) for mass, second (s) for time, Ampere (A) for electric current and Kelvin (K) for temperature. Decimal separator shall be ".", no thousand separator shall be used. If necessary, any conversion will be done in the selecting and editing step, before transferring the data into the repository.

All project members of PRESLHY will use standard terminology provided on the project website[4].

5 Increase Data Re-use

All published data is deemed to be re-usable. For the published data maintenance and accessibility of at least 25 years is envisaged, i.e. until end of 2045. However, the currently most probably selected repository will guarantee persistence of data for only 10 years. In the course of the project an extension to the 25 years will be elicited.

There is no embargo foreseen and the license model will be the **Free Culture License Creative Commons 4.0 International**, with reference to the name. This shall make the published data attractive for re-use.

Data quality assurance is included in the review, selection and editing processes explained above. No reviews are planned after the end of the project (31 December 2020).

Another way to increase re-use of the data is to organise model or code benchmarking exercises during or after the project term and to advertise, respectively refer to the data in all related publications and dissemination events.

6 Allocation of Resources and Data Security

Responsible for data management in PRESLHY project is the project coordinator, represented by Dr.-Ing. Thomas Jordan, KIT[5]. A small **data management team** is formed with at least one representative from the experiments performing partners: INERIS, HSL and Pro-Science (executing the KIT experimental program). The data management team members are responsible for the respective local storage of the raw data and for the preparation of the metadata.

4 https://www.preslhy.eu/glossary/

[5] Email: thomas.jordan@kit.edu

Commonly the team will chose the long term repository system. Currently, there are two possible candidates for long-term data storage: **KITopen research data repository**[6] and the **RADAR system repository**.

Both repositories offer the following data storage and security services:

- Data Archival with Publication
- Sufficient retention period (> 10 years)
- Optional embargos
- DataCite DOI Interface for peer review
- Choice of licenses, no CC0 required
- Metadata indexed, e.g. by RADAR, DataCite, Clarivate, Google Source

Equally they provide data security by the following measures.

- For each dataset, a checksum is calculated upon ingest
- The checksum is re-calculated after writing to tape ('read after write')
- Copies of the data are stored at different locations
- Different hardware, software and administration
- Routine migration of data to new storage media, including fixity check

With the RADAR services the costs for 25 years and the estimated 2 TB storage will amount to[7]:

Service:
25 x 500 Euro = 12500 Euro net

+ per 1 Gigabyte published data

2000 GB x 6,37 = 12740 Euro net

Total: 25240 Euro net

The KITopen services are free. However, the term of services are initially limited to 10 years. Figure 6 shows the data input interface, which is only accessible for KIT staff.

After comparing the proposed options, a choice will be made by the data management team and the expected costs of services will be fixed. The management work package WP1 is deemed to have the required budget.

[6] www.bibliothek.kit.edu/cms/kitopen-workflow.php

[7] https://www.radar-projekt.org/display/RE/2016/05/09/Updated+price+information

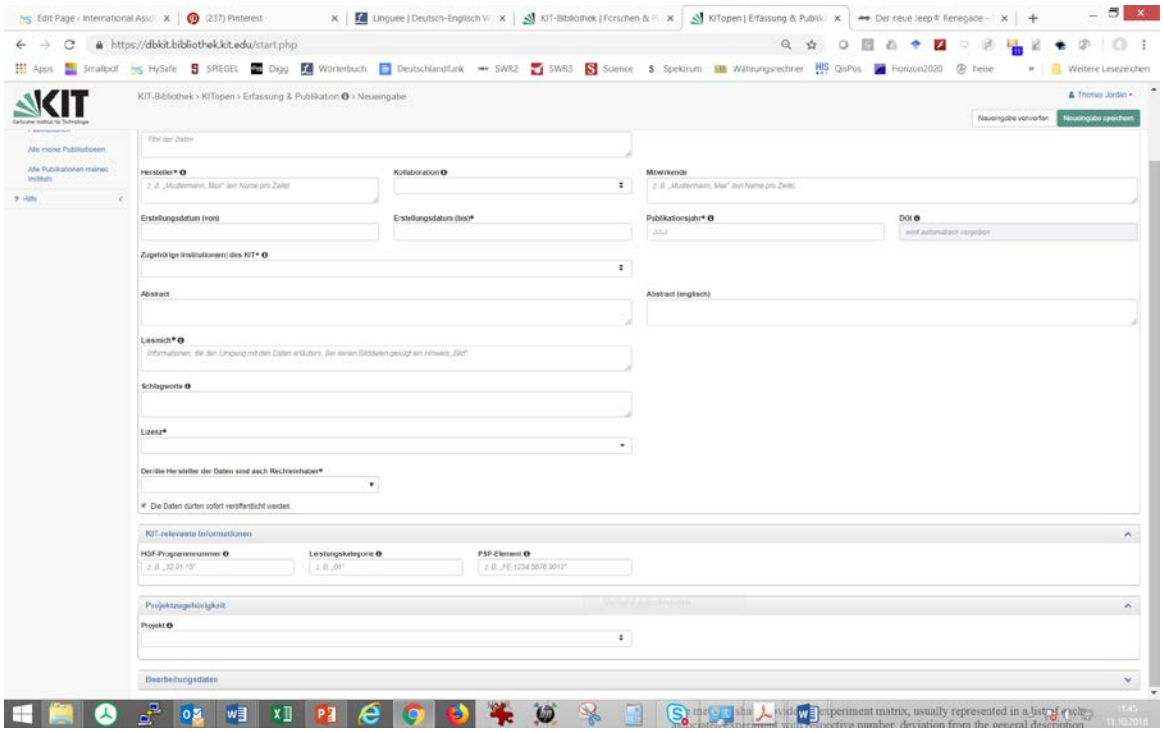

Fig. 6: Data input interface of KITopen