

# D1.3 Data Management Plan

**V0.3**

*Grant agreement: No. 101069689*

*From: Fraunhofer ISE*

*Prepared by: Sebastian Helmling*

*Date: 31/03/2023*

# TABLE OF CONTENTS

---

TABLE OF CONTENTS .....	2
ABBREVIATIONS.....	3
LIST OF FIGURES.....	4
VERSIONS.....	4
1. INTRODUCTION .....	5
1.1 Objective of the deliverable.....	5
1.3 Contribution of partners .....	5
1.4 Relation with other activities in the project .....	6
1.5 The FAIR data management principles.....	6
1.6 information exchange platform.....	7
1.7 protection of personal data.....	8
1.9 public deliverables .....	8
2. Monitoring .....	8
2.1 definition of system boundaries.....	8
2.2 Data collection .....	9
3. Data Storage and Processing Platforms.....	9
3.1. Locally at demonstration Sites.....	10
3.2. Cloud Integration of data sets.....	10
3.2.1. IoT-Platform.....	10
3.2.2. Backup strategy.....	11
3.2.3. Time Series Storage.....	11
3.2.4. Meta data .....	11
3.2.5. Data storage Protection.....	12
3.2.6. Data access for measurement data and derived data.....	13

3.2.7. Data Destruction.....	14
4. Data Protection Impact Assessment.....	14
4.1 Risks.....	14
4.2 Measures to reduce the risks.....	15
5. List of references.....	16

## ABBREVIATIONS

---

ACL	Access Control List
DPIA	Data Protection Impact Assessment
GDPR	General Data Protection Regulation
HUT	Heat Upgrade Technology
IoT	Internet of Things
IPR	Intellectual Property
WP	Work Package

## LIST OF FIGURES

---

Figure 1: mondas IoT Platform.....	11
Figure 2: meta data management tool.....	12

## VERSIONS

---

No.	Name	Partner	Contribution	Date
0.1	Sebastian Helmling	Fraunhofer ISE	First complete draft	17/02/2023
0.2	Sebastian Helmling	ALL	Checked by the partners	08/03/2023
0.3	Sebastian Helmling	Fraunhofer ISE	Final update	31/03/2023

## 1. INTRODUCTION

---

# 1.1 Objective of the deliverable

The main objective of the deliverable is to provide a data management plan for the measurement data which is collected and the derived data which are generated within the PUSH2HEAT project.

The Data Management Plan is the central document describing the life cycle of the data from collection to archiving, including all measures to ensure that the data remains available, usable, and comprehensible (understandable).

During the project measurement data from the heat upgrade technologies (HUT) will be gathered at the three/four demonstration sites. This requires the installation of sensors and data acquisition systems on site. Furthermore, additional data sources such as the controllers of the heat upgrade devices and the process automation of the plants hosting the heat upgrade technologies must be connected.

This requires the gathering, extraction, processing, storage and management of data sets coming from the PUSH2HEAT demonstration sites. This document specifies the data management plan that will be employed by the PUSH2HEAT project to ensure that all relevant data management policies are adhered to and to ensure that reliable data can be extracted and used for the PUSH2HEAT project development.

This document is being developed in accordance with the H2020 Guidelines on Data Management [1].

This document is delivered in Month 6 of the project. It will be updated every 12 months and whenever it becomes necessary e.g. because of the integration of new data sets.

## 1.3 Contribution of partners

The task T1.3 in WPI is led by Fraunhofer ISE and participated by the WP Leaders, i.e. TECNALIA, FRAUNHOFER, POLIMI, OST and EHPA.

Since all project partners have an interaction with the measurement data, their input to this document is necessary.

## 1.4 Relation with other activities in the project

The data management plan will set the basis for the interaction with measured and derived data in the project. It defines the rules for the interaction of the stakeholders of the project. The following tasks have a relation to that data and are therefore related to this deliverable.:

- T3.5. Commissioning and first performance tests
- T3.6. Assessment of commissioned demonstration systems.
- T4.1. Monitoring plan
- T4.2. Monitoring system integration and validation
- T4.3. Monitoring, performance data analysis
- T5.2. Replication studies
- T5.3. Life cycle environmental and cost assessment (LCA/LCC)
- T6.3. Communication and dissemination activation: creating impact and expanding outreach.

## 1.5 The FAIR data management principles

The FAIR (Findable, Accessible, Interoperable, and Reusable) data management principles provide a framework for making research data more accessible and reusable. The principles are designed to ensure that data are available for reuse and that they are of high quality, reusable, and reliable. The Horizon Europe open access mandate aims to promote open access to scientific publications and research data, and it aligns with the FAIR data management principles. Here's a step-by-step breakdown of how PUSH2HEAT meets the FAIR data management principles [2].

1. Findable: The first principle of FAIR data management is that data should be findable. This means that data should be easily discoverable and accessible. To meet this principle, the Push2Heat research data will be deposited and indexed in the OpenAIRE [3] infrastructure.

2. Accessible: The second principle of FAIR data management is that data should be accessible. This means that data should be available for reuse without any legal, technical, or financial barriers. To achieve this goal, an attempt is made to make the Push2Heat results available if no trade secrets or personal rights are violated by doing so.
3. Interoperable: The third principle of FAIR data management is that data should be interoperable. This means that data should be structured in a way that allows it to be easily integrated with other data sources. To meet this principle, PUSH2HEAT uses a standardized data format and a metadata schema.
4. Reusable: The fourth principle of FAIR data management is that data should be reusable. This means that data should be of high quality and suitable for reuse in different contexts. To meet this principle, PUSH2HEAT provides the documentation of the metadata that describes research data.

## 1.6 information exchange platform

Push2Heat uses the Microsoft 365 SharePoint to manage all the documents in the project.

Microsoft 365 SharePoint is a cloud-based platform that allows the PUSH2HEAT partners to collaborate and share files and resources. It is part of the Microsoft 365 suite of applications and provides a range of tools for creating, managing, and sharing content, including documents, images, and videos. SharePoint is used to manage their PUSH2HEAT workflows and content in a centralized and secure way.

SharePoint provides powerful search capabilities, allowing PUSH2HEAT partners to quickly find the content they need. In addition, it includes version control and auditing features, which help ensure that content is always up-to-date and that changes are tracked.

# 1.7 protection of personal data

E-mail addresses and phone numbers of the PUSH2HEAT partners are personal information and covered by the GDPR [4].

All the contact data of the PUSH2HEAT partners is stored in the Push2Heat Microsoft 365 SharePoint which is password protected. This ensures that only authorized PUSH2HEAT partners have access to this data. It is not permitted to share this data without the permit of the respective partner.

# 1.9 public deliverables

The deliverables which are generated within the PUSH2HEAT will be accessible to the scientific and technological community, if not market sensitive. All the project papers will be published in full open access journals, as Open Research Europe, and stored in repositories that are accessible through the EC service OpenAIRE [2] and Google searches. Furthermore, the publications will be stored in the ZENODO [5] repository.

## 2. Monitoring

---

The heat upgrade systems will be examined under real operating conditions. Both an energy evaluation and an investigation of the operating behaviour are to be carried out. For this purpose, the systems are tested within the framework of a metrological examination.

# 2.1 definition of system boundaries

In order to be able to evaluate the systems based on uniform criteria, boundary conditions must be defined. These boundaries apply to all three/four sites in the same manner. First, the following three framework conditions are defined for the evaluation of the systems:



1. Heat upgrade technology (HUT) only. Purpose: calculate and monitor the performance of the machine under different operating conditions.
2. HUT + heat source circuits and heat sink circuits for evaluating the impact of the auxiliaries of the external circuits on the performances.
3. Entire energy system including other generators. Purpose: evaluate the absolute and relative impact of the HUT on the overall system (fuel savings, additional electrical consumption, both in specific and absolute terms).

## 2.2 Data collection

For evaluating the performance of the PUSH2HEAT technologies data must be collected. This data will be gathered from three main data sources:

Sensors, heat meters, electricity meters, etc. measuring the physical state variables among the heat upgrade devices. These sensors are connected to a local data acquisition system.

Controller of the heat upgrade device: The controller connects sensors whose measured values are of interest for the evaluation of the PUSH2HEAT system. These data sets are either transferred to the data acquisition system via an appropriate protocol or sent to the Fraunhofer ISE server via an appropriate protocol.

Production automation system: The process automation control system collects data that is of interest for evaluating the framework conditions of the technology. These data sets are either transferred to the data acquisition system via an appropriate protocol or sent to the Fraunhofer ISE server via an appropriate protocol.

## 3. Data Storage and Processing Platforms

---

The data collected as part of the PUSH2HEAT project may be stored, maintained, and processed at several different points depending firstly on the nature of the data and secondly the provision of this data for access to other PUSH2HEAT tools. The remaining of this section provides an outline of each option that has been defined.

## 3.1. Locally at demonstration Sites

For the validation and demonstration of the PUSH2HEAT solutions, some components will run locally at the pilot sites. This includes tools that are installed and operated on local embedded systems that will generate data. These data are classed as PUSH2HEAT system data, and the same management and approval process as other data sets will be applied. The site-specific storage, protection, retention, and destruction of this data will be defined during the integration specification and plan for each of the pilot sites. The data generated locally are time series data from the sensors, counters and energy meters on site.

## 3.2. Cloud Integration of data sets

### 3.2.1. IoT-Platform

The IoT-platform tool mondas is used as a central data storage and exchange platform. Mondas is a IoT-cloud platform that provides access to the functionality for integration, security, storage, rule-based evaluation, and visualisation of the data sets. The tool mondas is hosted and maintained on a Fraunhofer ISE server located in Germany.



Figure 1: mondas IoT Platform

## 3.2.2. Backup strategy

The measured raw data is stored in daily raw data files before it is integrated in the HDF5 data bases. Both the raw data files as well as the HDF5 database is secured by a triple backup strategy. This means that all data is stored in parallel on two independently operating backup systems. The backup is operated by cron jobs organizing the synchronization.

## 3.2.3. Time Series Storage

Time series data are either pushed to or pulled to the mondas IoT-Plattform running in the Fraunhofer ISE Server. The time series data are stored using a HDF5 database. The Hierarchical Data Format version 5 (HDF5) is an open-source file format that supports large, complex, heterogeneous data. One of the main benefits of HDF5 is providing a transparent and efficient way of storing and handling many data sets in a small number of files. Furthermore, HDF5 offers versatile means of storing measurement and metadata in the same file structure. The time step for the time series data processed and stored in the project will be 60 seconds.

## 3.2.4. Meta data

Metadata and meta definitions play a critical role in the PUSH2HEAT data management, as they provide essential information about the data that allow for its effective management

and use. The metadata used in the measuring network includes descriptive metadata, which describes the data, administrative metadata, which provides information about the data's ownership, rights, and access, and structural metadata, which defines the structure of the data.

Every datapoint has a set of metadata. The meta data definitions apply in the same manner for all three demonstration sites.

The meta data are managed with a meta data management tool which is included in the monads framework.

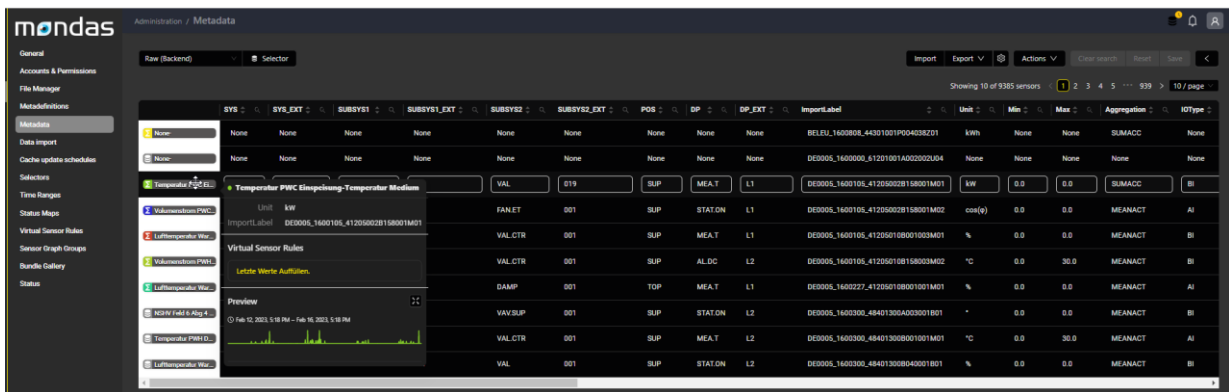


Figure 2: meta data management tool

## 3.2.5. Data storage Protection

Fraunhofer ISE provides a trusted and secure IT-platform for the PUSH2HEAT data which are generated and analysed during the project and after the project. The data transfer to the mondas IoT-Platform, the server which is hosting the platform and the web visualization are operated on Fraunhofer ISE servers. These servers meet ISO 27001 and ISO 27002 standards. Fraunhofer ISE provides functional, infrastructure, operational, and physical security for the IoT-Platform mondas.

## 3.2.6. Data access for measurement data and derived data

There will be a restricted web portal accessible via a member's section of the website <https://mondas.ise.fraunhofer.de/push2heat>. It will be used to give the project partners access to the measured raw data and to the derived data as result of the data analyzation processes.

The website is maintained by Fraunhofer ISE staff. It is hosted on a server located in Germany. These servers work on a Linux-enabled system, with back-ups provided by a large, experienced team of IT specialists offering the capacity to ensure rapid response, combined with professional, high-quality expertise.

In addition, the following definitions are utilized:

- Admin files are password protected
- Database is password protected

The Access Control List, or ACL, specifies which users or system processes are granted access to the data, as well as what operations may be performed on the data sets. Site administrators can control two separate features to the websites Access Control List:

- Defines which users can gain access to specific parts of the website. For example, will a given file be visible to a given user? A registered user will be able to view it, but the general public will not. The file might be invisible to all but a specific group of users, or further restricted to higher-level users.
- specifies what operations (or actions) can a user perform on any given data

For example, whether a user listed as an "Editor" has the authority to submit a file, or only to view it. The ACL settings can be set to permit submission and editing, modification of a file category, or any other combination.

Administrators can approve permissions for additional administrators, create un-limited additional group levels, assign users and define flexibility, as well as setting or modifying their authorizations:

- Access frontend
- Specify level of views which can be viewed
- Viewing data files

- Download data files

## 3.2.7. Data Destruction

The data remains in the full control of the owner, and all data generated by a partner can be deleted on request, with consideration of consortium agreement and the project exploitation and IPR requirements.

## 4. Data Protection Impact Assessment

---

Data Protection Impact Assessment (DPIA) is a process that aims to identify and minimize the privacy risks of the data generated, handled and stored and made public during the project and after its duration. It is an important tool for complying with data protection regulations, such as the General Data Protection Regulation (GDPR), and for ensuring that data processing activities do not infringe on the rights and freedoms of the project partners. The DPIA is important for PUSH2HEAT project because it involves the collection and processing of data related to industrial processes of companies. These data are sensitive, and their collection and processing must comply with data protection regulations, such as the GDPR.

### 4.1 Risks

A DPIA for identifies the following risks in relation to the data in the project:

- Data breaches: The data collected during the monitoring project may be targeted by cybercriminals seeking to steal sensitive information of the project partners. If the data is not properly secured, it could be accessed by unauthorized third parties.
- The data connections to the data acquisition system inside the company network and to the production automation can be a gateway for hacker attacks.
- Inaccurate data: The measurement instruments used to collect data may not be calibrated correctly or may be damaged, leading to inaccurate data collection.
- Data manipulation: If the data are not properly secured, they could be altered by unauthorized individuals to provide false readings, which could lead to incorrect conclusions being drawn from the data.
- Loss of data: If the data are not backed up or stored securely, it could be lost due to hardware failure or accidental deletion.

- Data subject privacy: The data collected during the monitoring project may contain personal information about employees or contractors involved in the industrial process, which could be misused or shared without their consent.
- risk of using the data for purposes other than those for which it was collected

## 4.2 Measures to reduce the risks

- The collection and storage of the data is done on servers of Fraunhofer ISE. These servers apply to the data security policy of the Fraunhofer Gesellschaft.
- The data transmission from the demonstration sites to the central server uses state of the art protocols which are permanently updated and follow the latest security regulations.
- Limited access to data: Access to the data collected by the monitoring system is limited to only those individuals who require it for the purposes of the project.
- Data minimization: Only data which is necessary to perform the evaluation of the HUT technology will be collected. This means avoiding the collection of unnecessary personal data or data from the production process.
- Where possible, data will be anonymized to prevent the identification of individuals.
- Data retention: Data collected by the monitoring system will only be retained for as long as necessary and should be securely deleted when no longer needed.

## 5. List of references

---

- [1] [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- [2] <https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-publications>
- [3] <https://provide.openaire.eu/home>
- [4] <https://gdpr.eu/>
- [5] <https://zenodo.org/>