D1.3

Data Management Plan – 2nd version

31st December 2021

OPTIMAI

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 958264

DOCUMENT SUMMARY INFORMATION

Grant Agreement No	958264	**Acronym**	OPTIMAI
Full Title	Optimizing Manufacturing Processes through Artificial Intelligence and Virtualization		
Start Date	01/01/2021	**Duration**	36 months
Deliverable	D1.3: Data Management Plan – 2nd version		
Work Package	WP1 – Project Management		
Nature	ORDP: Open Research Data Pilot	**Dissemination Level**	Public
Lead Beneficiary	ENG		
Authors	Cinzia Rubattino (ENG) Sabrina Verardi (ENG)		
Co-authors	Agata Gurzawska (TRI) Paul Hayes (TRI) Emma Teodoro (UAB) Andrea Guillén (UAB)		

DISCLAIMER

The OPTIMAI Project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 958264. The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the European Union. The European Commission is not responsible for any use that may be made of the information contained herein.

DOCUMENT HISTORY

Version	Date	Changes	Contributor(s)
V0.1	15/11/2021	Updated ToC	Cinzia Rubattino (ENG) Sabrina Verardi (ENG)
V0.2	03/12/2021	Review of contents provided in section 2 and 3	Paul Hayes (TRI)
V0.3	10/12/2021	New sections 3.6, 3.7, 3.8, 3.9 Updates for sections 1, 3 Datasets updated in section 4	Cinzia Rubattino (ENG) Sabrina Verardi (ENG)
V0.4	14/12/2021	Internal Review Version	Cinzia Rubattino (ENG) Sabrina Verardi (ENG)
V0.5	15/12/2021	1st version with suggested revisions	Paul Hayes (TRI)
V0.6	16/12/2021	2nd version with suggested revisions	Andrea Gomez (UNIMET)
V0.7	22/12/2021	Revisions after internal review and quality check	Cinzia Rubattino (ENG) Sabrina Verardi (ENG)
V1.0	23/12/2021	Final version for submission	Cinzia Rubattino (ENG) Sabrina Verardi (ENG)

PROJECT PARTNERS

Logo	Partner	Country	Short name
	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	Greece	**CERTH**
	FINT FUTURE INTELLINGENCE LIMITED	Cyprus	**FINT**
	IDRYMA TECHNOLOGIAS KAI EREVNAS	Greece	**FORTH**
	EVT EYE VISION TECHNOLOGY GMBH	Germany	**EVT**
	VISUAL COMPONENTS OY	Finland	**VIS**
	YOUBIQUO SRL	Italy	**YBQ**
	PANEPISTIMIO THESSALIAS	Greece	**UTH**
	ENGINEERING – INGEGNERIA INFORMATICA SPA	Italy	**ENG**
	UNIMETRIK SA	Spain	**UNIMET**
	UNIVERSITAT POLITECNICA DE VALENCIA	Spain	**UPV**
	Carr Communications Limited	Ireland	**CARR**
	UNIVERSIDAD AUTONOMA DE BARCELONA	Spain	**UAB**
	TRILATERAL RESEARCH LIMITED	Ireland	**TRI**
	KLEEMANN HELLAS –INDUSTRIAL COMMERCIAL SOCIETE ANONYME FOR MECHANICAL CONSTRUCTION SA	GREECE	**KLEE**
	TELEVES SA	Spain	**TVES**
	MICROCHIP TECHNOLOGY CALDICOT LIMITED	United Kingdom	**MTCL**

LIST OF ABBREVIATIONS

Abbreviation	Definition
AI	Artificial Intelligence
DMP	Data Management Plan
DOI	Digital Object Identifier
DT	Digital Twin
DX.X	Deliverable
EB	Ethics Board
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
GA	Grant Agreement
GDPR	General Data Protection Regulation
LIA	Legitimate Interest Assessment
N/A	Not Available
ORD	Open Research Data
R&I	Research and Innovation
UK	United Kingdom
WP(L)	Work Package (Leader)

Executive Summary

This report is the second version of the Data Management Plan (DMP) for the OPTIMAI project, provided as D1.3 - Data Management Plan – 2nd version in month 12 of the project. The overall purpose of this document is to support the data management lifecycle for all data that will be collected, stored, processed or generated by the project in order to maximise its access, according to the H2020 Pilot on Open Research Data (ORDP) in which the project participates.

The DMP aims to identify the scope for data management within the project and then to consider in turn the datasets present within the project.

The OPTIMAI approach will be in full compliance with the EU legislative and regulatory framework for ethics and data protection. So, in this document main regulations and basic concepts of the EU legal framework are summarised. A Data Management Plan Methodology is defined to provide the general rules and mechanisms for the access management of project data. Each dataset in the project is identified and described and information is provided about the extent to which it is standard compliant, and how the data will be available, accessible, interoperable and reusable.

The approach to data management within OPTIMAI is presented by establishing the types of data likely to be encountered, the FAIR approach to data management and how it is specifically applied within the project, processes for the management of personal data and further ethical and security considerations. The complete legal and ethical framework of OPTIMAI will be developed and reported in WP9.

The final part of the report then reviews each work package and task for any related data management requirements identified until M12, establishing an initial list of datasets within the project. So far, a total of 22 datasets have been identified in this version of the deliverable, being most of them data that will be collected and processed by the four pilot use cases. Other datasets are mostly stakeholders' information and public deliverables produced within the project scope. Although at this stage, some of this information remains incomplete or is yet to be determined this overview provides a solid basis for the data management lifecycle.

Overall, the OPTIMAI data management plan is intended to be a living document that is regularly updated throughout the remainder of the project as information about the data present within the project becomes more complete and new datasets emerge. As living document, it will be updated during the course of the project in new versions at M24 and M36.

Table of Contents

Executive Summary 6
Table of Contents 7
1 Introduction 11
1.1 Relation to other project work 12
1.2 Structure of the document 12
1.3 Summary of changes 13
2 Legal framework 14
2.1 European Convention on Human Rights 14
2.2 Charter of Fundamental Rights of the European Union 15
2.3 General Data Protection Regulation (Regulation (EU) 2016/679 of 27 April 2016) 15
3 Data Management Plan Methodology 17
3.1 Data Management lifecycle in OPTIMAI 17
3.1.1 Data types 18
3.1.2 Data collection 19
3.2 FAIR principles 22
3.2.1 Making data findable 23
3.2.2 Making data accessible 25
3.2.3 Making data interoperable 27
3.2.4 Making data reusable 27
3.3 Management of personal data 28
3.3.1 Personal data 28
3.3.2 Process for data protection in OPTIMAI 29
3.3.3 Purposes and Legal Bases of Processing Personal Data [TRI-UAB] 30
3.3.4 Data minimisation, storage, and retention 31
3.3.5 Rights of individuals in relation to their personal data and how they can assert them 31
3.3.6 Controllership and Responsibility for Data Management 33
3.4 Responsible Research & Ethics 34
3.4.1 Ethics research framework 34
3.4.2 Legal and Ethical Work Package (WP9) 35
3.4.3 OPTIMAI Ethics Board 36

3.5 Security considerations 37
3.6 Data Storage and Backup 39
3.6.1 Storage 39
3.6.2 Backup and Recovery 40
3.7 Data archiving and preservation 40
3.7.1 Short Term 40
3.7.2 Long Term 40
3.7.3 Data deletion 40
3.8 Allocation of resources 41
3.8.1 Costs for making data FAIR 41
3.8.2 Responsibilities of partners 42
3.8.3 Costs of potential value of long-term preservation 42
4 OPTIMAI Datasets Update 43
4.1 WP1 – Project Management 43
4.2 WP2 - User requirements, Technical Specifications and Use case analysis 46
4.3 WP3- Adaptive sensorial network for product-process monitoring and analysis of defects in manufacturing 46
4.4 WP4 - Production virtualization 58
4.5 WP5 - Novel techniques and systems for fast production (re)-configuration and planning 61
4.6 WP6 - Decision support and system integration 67
4.7 WP7 - Industrial demonstrator 69
4.8 WP8 - Dissemination, commercialization and exploitation strategies 70
4.9 WP9 - Legal and Ethical framework 73
5 Conclusions 75
References 76
Annex I Informed Consent forms template to be used at internal workshops and similar events 77
Annex II Data Minimisation Explainer 80
Annex III Dataset template 83
Annex IV Personal data template 85

LIST OF FIGURES

Figure 1: FAIR Data principles 23
Figure 2: Approach for Open Access research data (H2020 Online Manual) 26

LIST OF TABLES

Table 1: Fields to define datasets in the OPTIMAI DMP 19
Table 2: Fields to define Personal Data in OPTIMAI DMP 21
Table 3: WP9 Deliverables.....36
Table 4: Overview of tasks and datasets in WP1.....43
Table 5: Overview of tasks and datasets in WP2.....46
Table 6: Overview of tasks and datasets in WP3.....46
Table 7: Overview of tasks and datasets in WP4.....58
Table 8: Overview of tasks and datasets in WP5.....61
Table 9: Overview of tasks and datasets in WP6.....67
Table 10: Overview of tasks and datasets in WP7 69
Table 11: Overview of tasks and datasets in WP8 71
Table 12: Overview of tasks and datasets in WP9 73

1 Introduction

Industry is the backbone of the European economy. Enabling technologies like artificial intelligence (AI) and digital twins (DT) are powering remarkable growth potential and driving the next generation of industry. Using these technologies, the EU-funded OPTIMAI project will strive to strike an optimal balance between fast, cheap, and reliable production choices that have a big impact on the competitiveness of an industry. To this end, the project will develop smart instrumentation of production with AI-enabled sensors for quality inspection and monitoring. Using AI models, collected data will be analysed for the early detection of defects, the identification of upstream deficiencies and the reconfiguration of production parameters. To optimise production planning, the project will virtualise production using digital twins of processes and sensors.

Given the scope of the project, data will play a crucial role in all aspects of research, development, and evaluation of the OPTIMAI.

Depending on the domain, the data could derive from various sources such as laboratory testing, field trials, social science research and various observations. One of the major problems in these kinds of projects is the uncertainty of what will happen to the data after it is analysed, and the project has finished. In fact, the majority of data created can be of high value for other researchers, but because it is either stored on local servers and/or missing crucial metadata, its potential value is lost.

In this context, all partners of OPTIMAI's consortium adhere to sound data management principles in order to ensure that the meaningful data collected, processed and/or generated throughout the duration of the project is well-managed, archived and preserved, in line with the Guidelines on Data Management in Horizon 2020[1]. OPTIMAI is also participating in the Open Research Data Pilot. The purpose of the OPTIMAI Data management Plan (DMP) is to support the data management life cycle for all data that will be collected, processed or generated by the project according to the Guidelines on FAIR Data Management in Horizon 2020 [1]. Therefore, a comprehensive DMP is essential and must be delivered early in the project, fully describing the procedures for ensuring that the data management process complies with National (Greece, UK and Spain) and EU Legislation.

It is a document that outlines how research data will be handled during a research project, so it will be updated during the project's lifetime, and even after the project is completed, describing which data will be collected, processed or generated, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

Along these lines, this DMP aims to achieve the following objectives:

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

- Describe the data management lifecycle for the data to be collected and/or generated in the framework of OPTIMAI, serving as the key element of good data management.
- Outline the methodology employed to safeguard the sound management of the data collected, and/or generated as well as to make them Findable, Accessible, Interoperable and Re-usable (FAIR).
- Provide information on the data that will be collected and/or generated and the way in which it will be handled during and after the end of the project along with the standards applied to this end.
- Describe details on how the data will be made openly accessible and searchable to interested stakeholders as well as its curation and preservation.
- Present information on considering data security across the entire lifetime of data.

This deliverable focuses on providing a clear overview of the established data management practices within the OPTIMAI project and compiling the initial list of datasets present (or expected to be) within the project until M12. OPTIMAI will take a FAIR approach to data management following the European Commission's guidelines. Combined these two aspects comprise the overall OPTIMAI Data Management Plan.

Within this framework, we note, however, that data management is not a static process or a single activity but a practice that needs to be applied and refined throughout the project. Thus, this deliverable should act also as a living document capturing the details of new datasets as they emerge and updating the details of the ones already identified clarifying metadata, storage, archiving and access procedures as needed.

1.1 Relation to other project work

This DMP has impact on the actions that will be performed over datasets that will be collected, processed or generated. Specific impact is expected over the data collected and generated in WP1 Project Management, over data managed by the pilots in WP7, and finally over WP8 for the dissemination, communication and networking activities.

Moreover, this DMP is oriented to:

- The consortium partners.
- All stakeholders involved in the project.
- The European Commission.

1.2 Structure of the document

The remainder of the deliverable is structured in the following way:

- Section 2 describes the European legislative and regulatory framework for data protection, to which OPTIMAI should be compliant.
- Section 3 sets out the methodology for managing data within OPTIMAI including the types of data, the process of FAIR data management, the management of personal data, legal, ethical, and security considerations, the data storage and back policies, the guidelines for

the short- and long-term data archiving and preservation, the resources related to data management activities, roles and responsibilities for data management.
- Section 4 reviews the current status of existing and known future or expected datasets that will be present in each work package.
- Section 5 concludes the deliverable and sets out the path forward for managing data in the remainder of the project.

1.3 Summary of changes

This deliverable D1.3 is the second release of the OPTIMAI Data Management Plan (DMP) which updates and replaces the previous one (D1.2). Main changes for this release concern the update of the datasets in terms of the addition of new ones as well as the update and integration of previous ones. Accordingly, the whole data management strategy has been revised and updated where necessary, in order to ensure that OPTIMAI data handling is in line both with the project evolution and its results and with the EU guidelines about data management.

This section highlights the updates made in this second version D1.3 at M12 with regards to first version D1.2 delivered at M6:

- Updated Executive Summary to adapt the contents to the current version.
- Minor updates in Section 1 for the introduction.
- Addition of sub-sections 1.1 and 1.2 to better define DMP relevance to the project and structure.
- Addition of this sub-section 1.3 (Summary of Changes).
- Minor updates and corrections in sub-section 3.3.5.
- Minor updates in sub-section 3.5.
- Addition of sub-sections 3.6, 3.7, and 3.8.
- Updated Section 4 with revised and new OPTIMAI datasets.
- Updated Conclusions.

2 Legal framework

The OPTIMAI approach will be in full compliance with the EU legislative and regulatory framework for ethics and data protection. This chapter summarises the main regulations and basic concepts of the EU legal framework.

2.1 European Convention on Human Rights

The OPTIMAI project will adhere to the principles and spirit of the case law of the European Convention of Human Rights in the processing of any personal data that occurs during the project's lifecycle. The European Convention on Human Rights enshrines a non-autonomous right to data protection mainly through Article 8 right to respect for private and family life, which states:

1. Everyone has the right to respect for his private and family life, his home and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Additionally, the right to data protection also arises from Article 9 freedom of thought, conscience and religion; Article 10 freedom of expression; Article 14 prohibition of discrimination; Article 1 of Protocol No. 1 right to peaceful enjoyment of possessions; and Article 2 Of Protocol No. 4 freedom of movement (Council of Europe/European Court of Human Rights 2021) [6].

Following the relevant case law of the Court with respect to interferences with Article 8 rights, where personal data is processed it will be:

- Minimised only to what is legitimate and necessary in achieving the project's goals
- Kept accurate and up to date
- Retained for no longer than necessary to achieve the project's goals (including a five-year period after the project has ceased where such data is required for any audits initiated by the European Commission)
- Limited only to the purposes for which they were collected/processed
- Transparency of data processing procedures for access to personal data [6].

Particular attention will be paid to the case law of the Court with regard to data collection by employers in the workplace. No arbitrary or covert surveillance or audio/visual recording of employees or their activities and communications (especially those unrelated to their work) will be undertaken during the course of the project, and consent will be obtained, as required, by employees in order to process their personal data [6].

Furthermore, understanding that case law of the Court has enshrined particular rights of data subjects, the OPTIMAI project will uphold these rights by:

- Providing data subjects access to their personal data
- Allowing them to amend or rectify any of their personal data
- Deleting their data upon their request where this is possible [6]

Such rights will be facilitated by the provision of contact details of data controllers where research data collection is taking place, to data subjects, and a central point of contact on the project's website.

The right to data protection and arising obligations for data controllers are enshrined in the General Data Protection Regulation and will be further detailed in that section.

2.2 Charter of Fundamental Rights of the European Union

The Charter of Fundamental Rights of the European Union specifically enshrines the right to data protection in Article 8 protection of personal data, which states:

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

The relevant ends of data protection are similarly enshrined in Article 1 **human dignity** and Article 7 **right to private and family life**.

The OPTIMAI project will upholds these rights during data processing activities over the lifecycle of the project by adhering to the rules and principles of the General Data Protection Regulation as well as adhering to the principles and spirit of the case law of the European Convention on Human Rights.

2.3 General Data Protection Regulation (Regulation (EU) 2016/679 of 27 April 2016)

All OPTIMAI projects partners will adhere to the requirements of the General Data Protection Regulation while processing personal data.

Personal data is defined in the GDPR Article 4(1) as:

(1) 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person

Whilst processing is defined in the GDPR Article 4(2) as:

(2) ‘processing’ means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

3 Data Management Plan Methodology

Data management, and specifically research data management, within OPTIMAI will follow the guidance provided by the European Commission in their Guidelines on FAIR data management in Horizon 2020 [1]. The data management plan (DMP) has two core aims within the project:

(1) To set out the overall guidance for the management of data within OPTIMAI.

(2) To monitor the datasets, present within OPTIMAI and ensure they are managed FAIR-ly.

This approach is taken to ensure consistency in data management within OPTIMAI and to support maximum use and re-use of datasets through the implementation of appropriate data management practices.

As a project, OPTIMAI participates in the Open Research Data (ORD) Pilot, so it is committed to following best practices for overall data management and aligns itself within the guiding principle of the ORD Pilot that data should be "as open as possible, as closed as necessary"[2]. If anything, the necessary security requirements imposed upon the project should motivate a high standard of data management practice to ensure that all data is appropriately recorded, documented, stored, and disposed of as necessary and in line with legal, ethical, data privacy and security requirements.

In this section, we set out the overall guiding principles for data management within the OPTIMAI project. Firstly, by considering broadly the purpose and type of data likely to be present over the duration of the project. Next, we set out the FAIR data management principles focusing on how each general principle is specifically applicable in OPTIMAI and any specific processes and requirements to consider. Then, we focus on personal data, which require special attention in data management. Finally, we highlight the key legal, ethical, data protection and security considerations for the management of data.

3.1 Data Management lifecycle in OPTIMAI

Data is the lifeblood of the OPTIMAI project. The very essence of the project is to create an industry ecosystem that will optimize production through Smart Instrumentation, Metrology, Artificial Intelligence, Virtualization and Augmented Reality. Therefore, it is important for the project to track and understand the data that is used at each stage of the project.

The first stage of the data management process is to understand the broad context of the dataset in question; it is only then that the process of making the data FAIR can begin. Identifying this initial context of the dataset can be summarised by answering the following six questions as set out in the European Commission's (2016) guidelines:

[2] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

1. What is the purpose of the data collection/generation and its relation to the objectives of the project?
2. What types and formats of data will the project generate/collect?
3. Will you re-use any existing data and how?
4. What is the origin of the data?
5. What is the expected size of the data?
6. To whom might it be useful ('data utility')?

The goal of this initial DMP should be, as a minimum requirement, to arrive to a data summary for each currently known dataset present within the project. Future versions of the DMP will then expand upon these details including the FAIR-ification process (i.e., the process that aims at addressing the translation of raw datasets into FAIR datasets in a general way) as described in Section 3.2 below.

Therefore, a summarisation of a dataset within OPTIMAI should be able to set out the work package(s) and task(s) in which it appears, and how the dataset is related to the goals, research/development activities or evaluation of the specific task and the wider project (in relation to point (1) above). Secondly, the type of data collected should be considered, these are discussed in more detail in Section 3.1.1 below, but, in particular, are important to understand whether the data is collected, consumed or forms part of an evaluation within the context of OPTIMAI; the formats and filetypes and whether the data stands alone or resides within the OPTIMAI system is also essential (this covers point (2)). Point (3) is particularly, but not exclusively, relevant to the machine and deep learning tasks within OPTIMAI given the emphasis on artificial intelligence (AI) and would include both existing openly available standard and benchmark research datasets as well as data provided by end users for the purposes of the OPTIMAI project and links directly to point (4) on the origin of the data. At this time, point (5) concerning the size of the data may be unknown or unquantifiable but could be considered as number of respondents to a survey, actual size in MBs or GBs, or number of features or labelled objects. Finally, the utility of the data should consider what the longer-term possibilities are for the data itself whether that be for the research team, industry, other H2020 projects or the wider manufacturing research community; this does, however, have inextricable links to security restrictions required within the project.

Nevertheless, a core data summary that considers efficiently the above points provides an important starting point for proper data management practices.

3.1.1 Data types

As mentioned in the previous section, there is a core set of data types to be considered within the OPTIMAI project. Not all data within the project necessarily constitute research data; however, the DMP should make provision for the management of all data. Furthermore, the type of data being considered impacts upon its purpose, use and availability in terms of long-term data management practices. At this stage in OPTIMAI, we consider that there are six main data types that can be used as umbrella terms within the project.

- **Project management data** – these datasets are maintained to support the administration, management and dissemination activities of the OPTIMAI project; they are unlikely to be needed to be made available for any wider use.
- **Primary data collected by partner in OPTIMAI** – these are datasets collected by OPTIMAI partners for the purposes of carrying out a project task. This can, for example, be data relating to a specific software module, results of interviews or surveys, or part of an evaluation.
- **Secondary data (not publicly available)** – these previously collected datasets that have been provided to the OPTIMAI project that is not publicly available.
- **Input Data Stream** – Data collected from hardware devices, cameras, sensors etc.
- **Derived data** – these are datasets created from the output of processing activity by a specific software module in OPTIMAI.
- **Publicly available dataset** – these are datasets that are already available to researchers. They include, but are not necessarily limited to, training or benchmark data that support the development of machine learning/deep learning models.
- **Synthetic / generated data –** these are any datasets that are created for specific purposes within the project that do not contain 'real' data. Examples may include data resembling a specific manufacturing activity created for testing and demonstration purposes within the project.

Other types of data may emerge within the project, but these six types should cover the majority of datasets encountered within OPTIMAI. The DMP will be updated should any further data types be identified.

3.1.2 Data collection

This section presents the summary of datasets that will be collected and generated during the project, and that the partners have identified at this stage (detailed in Section 3.8). This summary will be refined, complemented and enlarged with new datasets as the project and the different tasks and activities make progress, and new versions will be included in the next planned reports.

To collect this information, a template as shown in Table 1 has been distributed among all partners, with the corresponding fields to have a clear vision of the datasets that will be processed during the project. As part of those fields, partners have been asked whether the dataset includes personal data, and in the case of a positive answer additional information has been required according to the template in Table 2.

Table 1: Fields to define datasets in the OPTIMAI DMP

Category	No	Field name	Description
Overview	1	DataSet ID	Identifier of the datasets

	2	Dataset Title	Internal name of the dataset
	3	Work Package	Work Package where the dataset in involved
	4	Task/Deliverable	Task/Deliverable where the dataset in involved
	5	Partner	Organization / person
	6	Data Type	Category of data
	7	Data Format	Indicate xls/csv/etc.
Details	8	Description of the dataset	General description of the dataset
	9	Data size	Indicate size and unit (TB, PB, etc …). Indicate dataset dimensions (nº of observations/rows and nº of features/columns)
	10	Status	Specify if it is already in place, established, or planned
	11	Use in OPTIMAI	How the data is/will be used in OPTIMAI
	12	Use beyond OPTIMAI	How the data could be useful to other researchers beyond OPTIMAI
Open Data	13	Is the data open?	Define: Yes – Public; Yes – but restricted access, No
	14	Explanation	Justify of open access decision
	15	Storage Location	Indicate where this dataset is / going to be stored: external repository, partner database, project database, etc. (if open, specify repository
	16	Who	Indicate who is responsible for storing the data
	17	Metadata	Make data findable including provisions for metadata and how is this managed
	18	How	Define how can the data be accessed (software, techniques)
	19	Increase data re-use	Define how and when will the data be made available for re-use
ETHICS and Protection	20	Personal Data	Indicate if the dataset includes personal data. If Yes, an additional template has to be completed

	21	Security Requirements	Indicate if there are specific security measures (both technical and organizational) to be considered regarding the dataset
	22	Comments	Any other explanation about the dataset

Table 2: Fields to define Personal Data in OPTIMAI DMP

No	Field name	Description
1	Workpackage/Task	Refer to a specific task or subtask here
2	DataSet ID	Identifier of the datasets
3	Types of personal data to be processed	Indicate the general type of data, e.g., business-related contacts, contact information related to dissemination, personal information from workshops, questionnaires, does the information falls under the category of special categories of data, etc.
4	Data Source	Note here whether the data will come directly from the data subject, or you will receive it from a data base, whether the data will be originated within OPTIMAI or priorly.
5	Purpose	Indicate why you need this specific type of information
6	Legal basis	Specify under which legal ground as outlined by Art. 6/ Art. 9 GDPR do you process the personal information
7	Data Minimisation	Specify how you guarantee that you're not going the collect more information than you need
8	Main controller(s)	Specify who determines why and how the data is processed
9	Processors involved	Indicate if you are involved in this activity on your own or you're working with other partners as well
10	Joint controllership	Indicate if you determine the purpose and the means for data processing with any other partner. If yes, with whom

11	Data Recipients	Indicate if this information going to be shared and with whom
12	Applicable safeguards	Indicate what technical and organisational measures are in place to ensure a high-level of protection to the personal data

According to that, in the Section 3.8 we follow a top-down approach in the description of how (technical and non-technical) WPs and pilots are managing and processing data in their respective activities.

3.2 FAIR principles

The principle of FAIR data management emerged in 2014[3] following the identification of the need to optimise the use of research data particularly with regard to supporting computational analysis. These principles were formalised by Wilkinson et al. [2] who defined the FAIR Guiding Principles for scientific data management and stewardship. These principles stated that data should be:

- Findable.
- Accessible.
- Interoperable.
- Reusable.

Wilkinson et al. were keen to stress that good data management is not intended to be a goal in and of itself but a means to support continued 'knowledge discovery and innovation'; a further goal could also be considered to support scientific reproducibility and replicability. Furthermore, an important distinction emphasised in the OpenAIRE guide to FAIR data[4] is that data can undergo the FAIRification process without the final outcome being that the data (or even the metadata) should become openly accessible. Instead, it is actually by following the FAIR process appropriately that a well-defined and reasoned justification for these decisions can be documented.

The European Commission [1] also advocates for the usage of FAIR principles for data management within its guidelines. Indeed, the European Commission has highlighted the importance of making the data produced by European-funded projects Findable, Accessible, Interoperable and Reusable, with a view to ensuring its sound management, as well as boosting the dissemination of relevant information and the easy exchange of data. Thus, European FAIR data approach implements standards and metadata to make data discoverable, specifying data sharing procedures and which data will be open, allowing data exchange via open repositories as well as facilitating the reusability of the data.

[3] https://www.force11.org/fairprinciples
[4] https://www.openaire.eu/how-to-make-your-data-fair

Each of the four pillars has specific principles that data and/or metadata should achieve to be FAIR data.

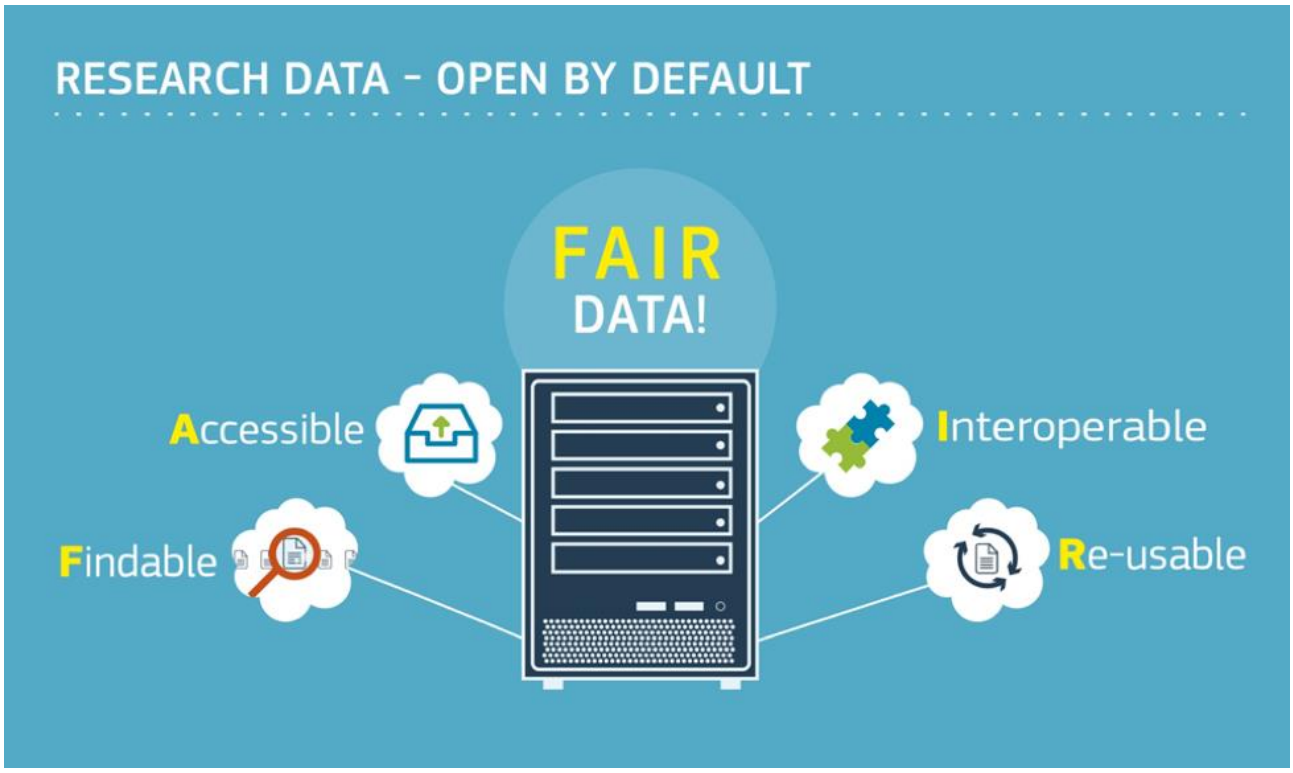

Figure 1: FAIR Data principles[5]

The living documentation for FAIR data is available at GO FAIR [3] but for clarity, the main principles of each pillar will be detailed within the subsection. In the sections below we set out what FAIR data means in the context of OPTIMAI by considering these guidelines, the original description from Wilkinson et al., and the GO FAIR Initiative.

This is intended to be living documentation and the DMP will be updated as the needs or requirements of the project evolve. Furthermore, as the datasets listed in Section 3.8 are realised the specific FAIR principles as applied to each dataset will be elaborated on.

3.2.1 Making data findable

Making data findable is the first core principle of the FAIR process. OPTIMAI emphasizes the need to improve the discoverability of data produced/used during its activities.

Regardless of the openness of the data, findability encompasses addressing the following factors. The GO FAIR sets four principles for helping to ensure data is findable.

3.2.1.1 Metadata

Appropriate use of **metadata** both through reuse of existing standards (such as FAIR sharing[6]) or clear processes when creating new metadata, use of **standard identifiers** (e.g., DOIs). Following a metadata-driven approach will improve the searchability of data, while at the same

[5] https://www.openaire.eu/how-to-make-your-data-fair
[6] https://fairsharing.org

time supporting its interpretation and re-use both for humans and computers. To this end, project data can be identified with rich metadata relevant to its content and format and 'machine-readable' to ensure automatic discovery of datasets and services. The project uses metadata that follow a globally unique and persistent identification mechanism for the development of rich and reliable metadata to promote the long-term discovery, usage and integrity of its data.

OPTIMAI is expected to deposit part of generated and collected data in an open online research data repository.

3.2.1.2 Naming conventions

Following common **naming conventions** – these should be consistent across the OPTIMAI project, the quality domain, and standard practices of industrial partners, where possible. Standards for the research area should also be observed. For the identification of OPTIMAI research data files it is recommended to use a descriptive name since its name will reflect the contents of the file and not use an exaggerated number of characters or special characters, and spaces.

The attributes to include in the file naming convention for OPTIMAI research data are presented in the following example:

'OPTIMAI_LCA_Inventory_Data_v0.1.xls'

The above syntax contains the following elements:

- A prefix to specify that it is OPTIMAI data (e.g., 'OPTIMAI').
- An intuitive title to the data (e.g., 'Dataset Example').
- For each new version of the data, specify the respective number (e.g., 'v0.1').
- The file respective format (e.g., 'xls').

3.2.1.3 Keywords

Similar to the above, **keywords** should also be defined and standardised, particularly in the case of open data. The project's data will be provided with easy-to-use search keywords in order to maximize its re-use by interested stakeholders throughout project's lifecycle. With that in mind keywords, as a subset of metadata, are used to add valuable information to the data collected/generated facilitating its discoverability and correlation to the OPTIMAI project.

In this regard, the project strategy on keywords is based on the following principles:

- The who, the what, the when, the where, and the why should be covered.
- Consistency among the different keyword tags needs to be ensured.
- Relevant, understandable and explicit keywording should be followed.

3.2.1.4 Versioning

Many datasets undergo changes as errata are fixed or new data is included, therefore **versioning** whenever a dataset is updated is essential.

Versioning makes a revision of datasets uniquely identifiable, thus, enabling us to keep track of the work done. More specifically, data versioning is used to define whether and how data changed over time, as well as to explicitly identify which version the creators / editors are working with. In addition, effective data versioning makes it easier to understand whether an updated version of a dataset is available and which changes are made between the different versions, allowing comparisons and avoiding confusion. In this context, a clear version number indicator is used in the naming convention of every data file produced during OPTIMAI in order to facilitate the identification of different versions.

3.2.2 Making data accessible

Making data accessible is the second core principle of the FAIR process. In OPTIMAI, accessibility is important given the nature of the industrial activities. There are broadly three main options when considering accessibility:

1. The data is not made available; however, a full justification should be included in the DMP and whether this applies only to external parties or also to members of the consortium. Reasons for non-availability may include legal, contractual, security, data privacy or intellectual property concerns.
2. The data is potentially available but is subject to a range of restrictions including but not limited to, who can use and access the data; how the data is accessed, the authorisations required; or how the data can be used.
3. The data is openly available on a research data repository and accompanied by comprehensive metadata and methodology of collection.

Should the decision be made that the data can be made available, this necessitates several further matters that should be considered to support the process.

- Where will the data be accessible from? (e.g., in a repository and which one?)
- Is there additional information that should be deposited with the data including the metadata, a codebook, software or source code?
- Does an authorisation process need to be in place for access and who manages this process? Is this process clear and transparent? Is it combined with an appropriate licence?

In order to maximize the impact of OPTIMAI project, research data will be 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. All data are made available for verification and re-use.

In OPTIMAI data available will make findable and accessible by providing a common repository for storing the data and offering a simple programming interface for accessing it.

Public deliverables listed in the GA will be made publicly available via the project website. All personal and sensitive information will be removed from these datasets/ reports before they are made public.

3.2.2.1 Research Data and Open Access

OPTIMAI project participates in the European Commission's Open Research Data Pilot (ORD) pilot, which applies mostly to the data needed to validate the results presented in scientific publications. Open access implies unrestricted online access to research outputs such as journal articles, without access fees as illustrated in Figure *2*.

The goal of the EU with this program is fostering access to and re-use of data generated by EU funded projects in order to improve and maximize public financial European resources and avoid duplication of efforts.

Participating in the ORD Pilot does not necessarily mean opening up all your research data. Rather, the ORD pilot follows the principle *'as open as possible, as closed as necessary.'*

The main aim of the H2020 Open Access mandate is to make research data generated in H2020 projects accessible, so it is possible to:

- build on previous results,
- encourage collaboration,
- speed up innovation, and
- involve citizens and society,

while also accepting restrictions and protection for data due to security and/or commercial reasons or privacy concerns (see Figure *2*).

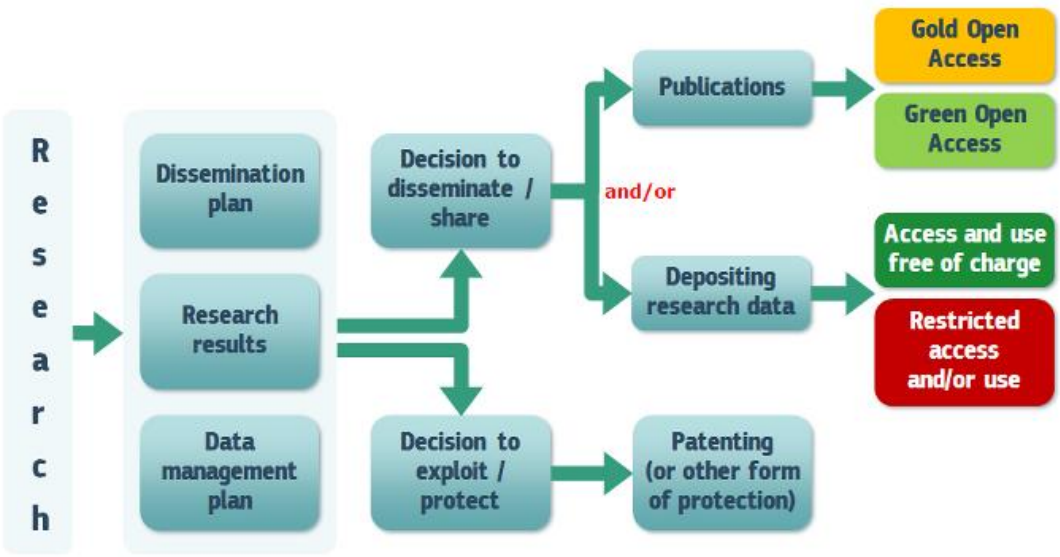

Figure 2: Approach for Open Access research data (H2020 Online Manual)[7]

In OPTIMAI, as included in the Grant Agreement, at a preliminary stage, partners agreed on open access publishing. Research publications will be made available for free access for everyone including the rights to read, make, download, print and right to copy, distribute, search, link, trace and extract. Open Access does not imply an obligation to publish results since this decision

[7] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/index_en.htm

is entirely the responsibility of the partners and does not also affect the decision to commercially exploit the results. Open access can only become a problem if the publication is chosen as the means of dissemination. The decision to publish (or not) through open access should only come after a more general decision on whether to publish directly or to first seek IP protection.

However, in the future, partners may also opt for gold or green access to peer-reviewed scientific publications, which might result from the project, depending on the type of information to be published.

3.2.3 Making data interoperable

Interoperability is the third core principle of the FAIR process. Datasets are often at their most valuable when combined with other data, facilitated by the interoperability process. Interoperability can be supported through:

- the use of standardised formats, compliance with existing standards, usage of common ontologies;
- the use of common metadata within the project.

Access to data is also a long-standing problem in industrial research activities, particularly in relation to the availability of domain specific datasets for the training and testing of machine learning and artificial intelligence algorithms.

It is essential that published datasets are unequivocally interpretable by third persons without any link to the project. Therefore, each dataset needs to be accompanied with a description of the methodology, sources, definitions, and scope of the data contained in it. Whenever possible, datasets should be structured in such a way that it can, in full or in part, be combined with another dataset, from the project or any other data source.

OPTIMAI will adopt in its data management methodology the use of metadata vocabularies, standards, and methods that will increase the interoperability of the data collected/generated through its activities.

More specifically, standard vocabularies will be used for all data types present in the project. In case there is an uncommon vocabulary, a clear mapping will be provided in order to facilitate its use. Thus, the project's data will be interoperable and easy for sharing among researchers, institutions and organisations.

Partners will observe OpenAIRE guidelines for online interoperability[8], furthermore they will also ensure that project data observes FAIR data principles[9].

3.2.4 Making data reusable

Reusability is the final core principle of the FAIR process. The promotion of reusability is important particularly when data is being made available to other researchers as the need to both understand the context in which the data was collected and any relevant limitations to the

[8] https://guidelines.openaire.eu/en/latest
[9] https://www.force11.org/group/fairgroup/fairprinciples

dataset as well as in which circumstances reuse is permitted. Reusability is focused, on the one hand, on the application of appropriate licences.

Whilst on the other it should also consider the timescales from collection to publication, the effect of any embargoes, or if there is a 'shelf-life' to the applicability of the data.

Other factors affecting reusability include to whom the data itself may be useful, e.g., data collected for use by a specific module, as part of a particular use case or for a single end user that may not transfer to other uses. Documentation accompanying the dataset submission may set out how data should (not) be interpreted to avoid misunderstandings.

OPTIMAI project is expected to produce a substantial volume of data and knowledge that will be presented to the scientific community, to the Digital Innovation Hubs community, industry, policymakers and society at large through a carefully designed portfolio of dissemination actions. Each dataset will have an individual license. Access to project data will be provided to the whole project Consortium and exclusively for the project objectives. Datasets produced as a result of the project work will be shared within the Consortium and will only be allowed for external sharing with a consensual Consortium approval of the relevant stakeholders, by accepting the terms and conditions of use, as appropriate. The license for the access, sharing and re-use of OPTIMAI material and output datasets will be defined by the Consortium on a case-by-case basis.

Finally, before any OPTIMAI data is made available **quality and security assurances** should be in place to both ensure that the dataset is free from errors, is appropriately documented and does not raise any security concerns.

3.3 Management of personal data

3.3.1 Personal data

The data management process must be particularly mindful of the requirements for the management of personal data. In the context of OPTIMAI, the prevailing framework in which personal data must be managed is the European Union General Data Protection Regulation (GDPR) [4].

Personal data refers to any data that related to an identified or identifiable living individual (a so-called 'data subject'). Examples of personal data under the GDPR given by the European Commission include, but are not limited to[10];

- a name and surname,
- a home address,
- an email address such as name.surname@company.com,
- an identification card number,
- location data (for example the location data function on a mobile phone),
- an Internet Protocol (IP) address,

[10] https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en

- a cookie ID,
- the advertising identifier of your phone,
- data held by a hospital or doctor, which could be a symbol that uniquely identifies a person.

The processing of such data must adhere to the highest ethical and legal standards. The consortium's approach will be in full compliance with the EU legislative and regulatory framework for data protection based on the uniform approach of the GDPR, and the national legislative and regulatory framework for data protection of each project member country.

3.3.2 Process for data protection in OPTIMAI

GDPR defines 'processing' as any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

Article 5 of the GDPR lays out the principles relating to the processing of personal data. These are known generally as the seven key principles and cover:

- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality
- Accountability

The management of any personal data within OPTIMAI should follow these principles from initial collection through to long-term storage and archiving (or disposal) throughout the project. In particular, partners must assure that they have established a lawful basis for any data they will process. The accepted lawful bases set out in Article 6 of the GDPR and are one of the following:

- (informed) consent,
- necessary to perform a contract,
- necessary for compliance with a legal obligation,
- vital interests,
- performance of a public task, or
- necessary for your legitimate interests.

OPTIMAI will rely on the processing of personal data for research purposes, especially concerning contact details and opinions of those participating in the interviews, workshops and validation tests. The project will collect and process data only if, and insofar as, it is necessary for its research. Interviewees and validation test participants will be asked for their signed informed consent and a clear description of the procedures that will be used for data control

and anonymisation will be provided. A detailed explanation of the approach followed in OPTIMAI project is reported in the following sections.

3.3.3 Purposes and Legal Bases of Processing Personal Data [TRI-UAB]

Over the course of the project, personal data will be collected only insofar as it is necessary for the successful completion of project objectives and goals including to support core research activities, dissemination and exploitation activities.

As such, data processing activities in the OPTIMAI project will fall under two categories 1) research and 2) communication (dissemination and exploitation of results).

The legal bases for data collection for research activities will primarily be consent and legitimate interest (where seeking consent would be inadequate to the processing activity).[11]

Where possible and as a matter of priority, research participants/data-subjects will be recruited on the basis of informed consent, having been supplied with information sheets and having been given the option of refusing participation, or withdrawing consent at any time, without any adverse or negative consequences to them, in line with the requirements of Article 7. Partners will provide information of the lawful basis of data collection to data subjects, as well as their rights relating to their personal data and how to exercise them. Signed consent forms will be retained by data controllers in order to furnish as required by relevant authorities.[12]

If it is necessary to re-purpose existing datasets, this will be done on the basis that scientific research is compatible with its original intended purpose and such that no separate legal basis from that which allowed its collection is required.[13] Appropriate technical and organisational safeguards will be deployed to protect personal data.[14]

Where consent cannot be obtained, a legitimate interest assessment (LIA) will be conducted by partners responsible for processing any such data. A template of this assessment is available on repository of the project.

Project partners do not currently envision processing any special categories of data,[15] however in the event that this is necessary it shall be done with the express consent of data subjects, or on the basis of legitimate interest for the purposes of scientific research with technical and organisational measures including pseudonymisation or anonymisation implemented.[16]

For communication activities, the legal basis will also be consent. Interested persons will have the ability to subscribe to the OPTIMAI mailing list or newsletter by using the project website and

[11] Article 6(1)(a)(f)
[12] Article 7(1)
[13] Recital 50
[14] Article 89(1)
[15] As enumerated in Article 9(1): data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation
[16] Article 9(2)(a)(h) and Article 89(1)

will be duly informed of how their personal data will be processed, as well as their rights and how to exercise them. A legitimate interest's assessment will be conducted ahead of sending any newsletter to persons to whom the partner's decide may be interested.[17] Such persons will be free at any time to opt out of communications through clearly indicated means.

The rights of all data subjects will be safeguarded through appropriate organisational and technical measures, including through techniques such as anonymisation and pseudonymization and in accordance with the requirements of data minimisation.

3.3.4 Data minimisation, storage, and retention

Personal data collection will be restricted to no more than what is strictly necessary to achieve the objectives and goals of the project, which is to say that only data that is adequate, relevant, and limited to what is needed for the completion of tasks.[18] Any personal data that is collected and which is not necessary for the completion of tasks will be destroyed as soon as task/activity will be completed.

Personal data will be stored only for as long as necessary to achieve the goals and objectives of the project,[19] and stored personal data will be reviewed by partners annually with a view to determining its ongoing relevance to current and future tasks. Personal data may be retained for up to five years in order to comply with auditing requirements, for example, personal data deemed necessary for the execution of the audit of a project partner by the European Commission.

3.3.5 Rights of individuals in relation to their personal data and how they can assert them

The GDPR recognises and enshrines a number of rights that can be exercised by the data subject and as such confer duties upon data controllers. OPTIMAI projects partners will execute their duties in adherence with the requirements of the GDPR.

The right to be informed: Article 13 of the GDPR requires data controllers to inform data subjects about the processing of their personal data at point of collection. Article 14 requires the controller to inform the data subject about the processing of their personal data where it was collected by an entity other than the controller. Data subjects will be informed, as required by the GDPR, at point of data collection, through informed consent forms or other means (e.g., website privacy policies) of the purpose of data processing, retention periods or criteria, and with whom the data will be shared. Where controllers use data that were collected by another entity, they will make reasonable efforts to contact the concerned data subjects and provide them with the required information. Appropriate safeguards will be implemented where this is not possible, and the use of any such data will be publicised on the project website.[20]

[17] Article 6(1)(f)
[18] Article 5(1)(c)
[19] Article 5(1)(e) and Recital (39)
[20] Article 14(5)(b)

The right of access: Article 15 of the GDPR establishes the right of data subjects to know if their personal data is being processed and grants them the right of access to any such data. This information will be provided subject to the requirements laid out in Article 12.

The right to rectification: Under the provisions of Article 16, data subjects have the right to request the correction of inaccurate personal data, or the completion of incomplete data. OPTIMAI project partners will take seriously their incumbent duty of provide data subjects sufficient access to identify errors and incomplete information and correct and complete it as necessary.

The right of erasure: Article 17 enshrines a right to erasure of personal data at the request of the concerned data subject. Project partners will endeavour to comply with deletion requests, including from OPTIMAI datasets or contact lists to the extent that such deletion requests do not unreasonably compromise the successful carrying out of the project's objectives and goals.[21]

The right to restrict processing: Under particular circumstances, the provisions of Article 18 grant data subjects the right to restrict processing of their personal data. In the event that OPTIMAI project partners receive any request from a data subject to the effect of invoking this right, they will adhere to the requirements of Article 18(1)(a)-(d) and store the relevant data until the issue is resolved, whereby the personal data will be destroyed or processed in a manner that the data subject can and does consent to.

The right to data portability: Article 20 (1) grants the data-subject the right to receive their personal data, upon request, in a "…structured, commonly used and machine-readable format" as well as the right to "…transmit those data to another controller without hindrance from the controller to which the personal data have been provided". OPTIMAI project partners will strive to accommodate related requests as required where, in particular, the data was processes on the basis of consent, contract, or by automated means as and if applicable.

The right to object: Article 21 grants data subjects the right to object to the processing of their data. OPTIMAI project partners will obtain consent from data-subjects in advance of data collection and data-subjects will be free to decline to provide consent or withdraw it at any time. Data-subjects will be provided with the ability to unsubscribe from project communications.

Rights in relation to automated decision-making and profiling: Article 22 grant that data-subjects have the right not to be subject to automated decision-making or profiling which creates legal or similar effects for the data subject. Data subjects can however consent to such automated profiling or decision-making.[22]

International data transfers: Most OPTIMAI partners are based in the EU, with one located in the United Kingdom (MTCL) and another sharing digital infrastructure with its United Kingdom organisation (TRI). Whilst the United Kingdom is no longer a member of the European Union, on June 28, 2021 the European Commission adopted two adequacy decisions for the UK and no

[21] Article 17(3)(d)
[22] Article 22(2)(c), GDPR

exceptional restrictions are likely to be placed on data transfers to the UK over the lifetime of the OPTIMAI project.[23]

The OPTIMAI project's primary storage solution is NextCloud, a cloud storage software platform provided by a German GDPR compliant company[2425], the repository is administrated and maintained by CERTH following the principles of the OPTIMAI data protection framework.

Storage solutions and other digital infrastructure hosted by international corporations including Microsoft and Google may be utilised to store and transmit personal data or to manage mailing lists. Data subjects will be informed about international data transfers at point of collection of personal data and will be given the option of providing or declining their consent to the use of cookies where the project website is concerned. International services used will be compliant with Article 46(2)(d), which is to say utilisation of these services will be contingent on the existence of standard data protection clauses in the service provider's terms of service.

Ethical and legal requirements are specified in deliverables D2.1 and D9.1, with further requirements and guidance to be presented in deliverables D9.2 (M12), D9.3 (M12) and D7.2 (M18).

3.3.6 Controllership and Responsibility for Data Management

The OPTIMAI consortium consists of 16 organisations, each of which are distinct legal entities. It is intended that each partner will act as the controller[26] of the personal data it collects and processes and will take necessary precautions in order to remain single controllers during the course of OPTIMAI research and personal data collection and processing activities. As a consequence, each partner will individually be responsible for their own compliance with data processing rules.

Whereby one partner can be considered to be directing another partner in its personal data collection and processing activities such that it is determining the **means and purposes[27]** of those activities, the consortium acknowledges that a controller-processor[28] relationship may

[24] https://nextcloud.com/gdpr/
[25] https://nextcloud.com/privacy/
[26] GDPR Article 4(7) defines controller as: 'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law
[27] How the data is collected, and why it is collected.
[28] GDPR Article 4(8) defines processor as:
(7) 'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law

exist between these partners during the execution of any related tasks. In such cases partners will establish appropriate data processing agreements.[29]

The consortium also acknowledges that on other occasions in the execution of a task **multiple partners** may determine the **means and purposes** of personal data collection and processing activities, thereby establishing joint controllerships for the execution of such tasks. Where such a relationship can be determined to be in place between partners, they will sign a binding document that outlines their responsibilities and the nature of controllership and relevant contacts shall be made transparent such as through the project website.[30]

Project partners will be referred to *Annex I: Flowchart for applying the concepts of controller, processor and joint controllers in practice*, of the European Data Protection Board's *Guidelines 07/2020 on the concepts of controller and processor in the GDPR* (European Data Protection Board 2020, 46–48) [7] in order to help them assess their controllership status in advance of data processing activities.

3.4 Responsible Research & Ethics

3.4.1 Ethics research framework

3.4.1.1 Research integrity

The OPTIMAI Consortium will comply with the highest ethical standards, principles and good practices of research ethics laid down in the European Code of Conduct for Research Integrity [5]. Therefore, all OPTIMAI research activities should be conducted in strict compliance with the general **principle of integrity**, which entails conducting research activities abiding with the highest ethical standards and minimising risks and harmful results or consequences [5]. **The integrity principle** should be complemented with the following ethical principles, which should also be observed by OPTIMAI researchers:

- **Reliability:** Ensure the quality of the design, the methodology, the analysis and the use of resources in the research.
- **Honesty:** Develop, undertake, review, report, and communicate the research in a transparent, fair, full and unbiased way.
- **Respect:** Respect research colleagues, research participants, society, and the environment.
- **Accountability:** Be accountable for publication, management and organisation, training activities, supervision, and for the wider impacts of the research.

3.4.1.2 Responsible Research Innovation

The OPTIMAI Consortium fully adheres to the EU Ethical Responsible Research and Innovation Framework (RRI).[31] The purpose of RRI is twofold: i) to anticipate and assess how technological

[29] Article 28(3)
[30] Article 26(1)
[31] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

developments could have an impact on society and the environment, and ii) to ensure that technological developments respond to and align with individual and societal values, needs and expectations.[32] The OPTIMAI Consortium should follow the principles of RRI[33], namely:

- **Diversity and inclusion** imply being sensitive to research biases, avoiding discrimination and stigmatisation and striving for representativeness and diversity.
- **Anticipation and reflection** mean the need to assess the purposes of the research, its benefits and risks, outcomes, potential unintended consequences and impact on individuals and the society and envisage possible to address them.
- **Openness and transparency** involve being open to society in a meaningful and honest way.
- **Responsiveness and adaptive change** entail adapting the research plans to changing social values, needs and expectation, as well as to emerging knowledge and new insights.

The OPTIMAI project acknowledges that responsibility applies to both the research and innovation (R&I) process as well as its outcomes. OPTIMAI's objective is to contribute to responsible, ethical and sustainable impact, both at the R&I process (internal impact) as well as solutions developed as its result (external impact).

3.4.2 Legal and Ethical Work Package (WP9)

WP9 is devoted to the ethics and legal aspects of the project. The purpose of WP9 is to ensure that all research activities conducted in OPTIMAI are legally compliant, ethically acceptable and socially desirable. To that end, four tasks have been foreseen: T9.1 focuses on identifying the ethical and legal framework applicable to the project; T9.2 will identify and analyse the ethical, legal and societal risks that the research activities may pose; T9.3 entails the definition and implementation of the OPTIMAI Regulatory Model; lastly, T9.4 involves the ethical and legal monitoring strategy from an internal and external perspective.

In this regard, the different stages of the strategy designed for the monitoring of the ethical and legal implications of the OPTIMAI research activities reflect strict compliance with the RRI framework given that such a strategy: i) ensures a **more broadly** ethical and legal voluntary **engagement of society**; ii) allows for the **anticipation and assessment** of potential risks that may jeopardise human rights; iii) relies, in a complementary manner, on the **independent assessment** of the OPTIMAI Ethics Board; and, iv) focuses the ongoing monitoring strategy on the **legal, ethical and societal dimensions** (including environment and sustainability) of the research activities.

The deliverables to be submitted within WP9 are listed in the Table below (See Table 3). The involvement of the OPTIMAI Consortium as a whole is key to put into practice the RRI principles and to develop a robust solution that is legally compliant, ethically acceptable and socially desirable.

[32] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation
[33] https://rri-tools.eu/en/about-rri

Table 3: WP9 Deliverables

Due Date	Deliverable Number	Title	Lead Beneficiary
M6	D9.1	Report on the OPTIMAI ethical and legal framework	UAB
M6	D9.5	Report on the OPTIMAI Regulatory Model - 1st version	UAB
M12	D9.2	Report on OPTIMAI ethical, legal and societal risks - 1st version	TRI
M12	D9.6	Report on the OPTIMAI Regulatory Model - 2nd version	UAB
M24	D9.3	Report on OPTIMAI ethical, legal and societal risks - 2nd version	TRI
M24	D9.7	Report on the OPTIMAI Regulatory Model - 3rd version	UAB
M36	D9.4	Report on OPTIMAI ethical, legal and societal risks - 3rd version	TRI
M36	D9.8	Report on the OPTIMAI Regulatory Model - 4th version	UAB
M36	D9.9	Report on ethical, legal and societal impact of OPTIMAI	TRI

3.4.3 OPTIMAI Ethics Board

The OPTIMAI Ethics Board (hereinafter referred as EB) was formally appointed in a virtual meeting held on 6 May 2021 and led by the ethical partner of the project (UAB). The composition of the EB reflects the internal and external dimensions of the ongoing monitoring strategy designed for the project with the aim of ensuring ethical and legal compliance. To that end the cross membership of the EB was agreed by the OPTIMAI Consortium as follows:

- Dr. Stavroula Ntoa, Dr. Dimosthenis Ioannidis and Dr. Cinzia Rubattino as internal highly qualified experts nominated by FORTH, CERTH and ENG respectively.
- Dr. Ugo Pagallo (University of Turin, Italy) -appointed as EB Chairman-, Dr. Lilian Mitrou (Aegean University, Greece), Dr. Marta Poblet (RMIT, Australia) and Dr. Louis de Koker (La Trobe University, Australia) as independent renowned international experts nominated by the UAB.

The general tasks of the EB are the following:

- Reviewing, reporting and advising the Consortium concerning ethical matters in accordance with the nature and the aim of the project.
- Providing independent advice and guidance to the Consortium regarding potential risks posed by the OPTIMAI technologies that may have negative impacts on individuals and society.
- Monitoring the project modus operandi with reference to the foreseen research activities.

- Monitoring the design, development and implementation of the OPTIMAI solution and providing mitigation measures to minimise the risks that this solution may pose from an ethical perspective.

Additionally, the EB will carry out the following **specific tasks**:

- To produce six semi-annual reports on OPTIMAI research activities.
- To present an annual report during the Plenary Board Meeting at the end of the 1st, 2nd and 3rd year of the project.
- To review all Deliverables that pose ethical or legal concerns. The EB will review all deliverables of WP9 and WP7 and deliverables of WP2 (D2.1; D2.2; D2.4; D.2.5 and D.2.6). The revision of other deliverables that may entail ethical or legal concerns will be determined according to the project's developments.
- To maintain an open line of communication so that any member of the Consortium can consult with them ethical concerns.
- To be physically present (at least one external and one internal member of the EB and the rest of members will attend remotely) at the Plenary Board Meetings of the project once the situation with the COVID-19 pandemic improves. Until then, all the EB meetings will be virtually held.

OPTIMA EB **Terms of reference and working procedures**:

- Principles governing the activity of the EB: All EB members shall perform their tasks in compliance with the principles of excellence, independence, accountability, transparency and confidentiality.
- Voting rules: the EB shall act by a majority of its members.
- Decision-making procedures: Decisions, opinions or actions taken by the EB within the project shall be reached by consensus. If such consensus cannot be reached, the EB will act on a majority basis. Nevertheless, minority or dissenting opinions shall always be included separately in the corresponding Semi-Annual Reports and properly communicated to the Consortium and the Project Officer.
- Efficient functioning of the EB: The external member acting as Chairman of the EB (Professor Ugo Pagallo) will be the main contact person of the EB. The ethical lead partner (UAB) will assist the EB and in particular Prof. Pagallo to manage any ethical concern posed by the Consortium partners through their corresponding task or work package leaders.

3.5 Security considerations

OPTIMAI is funded under the European Commission's Horizon 2020 "Nanotechnologies, Advanced Materials, Biotechnology and Advanced Manufacturing and Processing" programme under the call "H2020-DT-FOF-11-2020: Quality control in smart manufacturing". Many of the activities within OPTIMAI are reported in deliverables which have either been classed as public or confidential.

OPTIMAI adheres to security policies in order to maintain the integrity of data and to make sure that the data will not be accessible by unauthorized parties or susceptible to corruption of data.

For data processed in relation to deliverables with a confidential dissemination level, it is a decision of the relevant partner(s) or the consortium as a whole as to whether the data can potentially be made available beyond the beneficiaries concerned; nonetheless, it should still undergo the FAIR process to ensure it is documented appropriately. Decisions relating to the long-term use, storage and archiving of such data should be compliant with the OPTIMAI ethics and legal framework defined in WP9 (see section 3.7).

In all cases it is important to consider not only the data itself but also

- Potential impact should this data be combined with other datasets.
- Contents of the metadata that would need to be provided with the dataset.
- Documentation detailing how the data was collected / created.
- Information relating to the processes required for parsing the data.
- Inferences that could be made from the data not considered under its original purpose.

Particular care should be applied to any data collected during piloting and demonstration activities that may highlight how multiple datasets can be linked together or could expose industrial analytical or investigative capabilities or processes.

OPTIMAI will securely handle any collected/generated data throughout its entire lifecycle as it is essential to safeguard this data against accidental loss and/or unauthorized manipulation. Particularly, in case of personal data collection/generation it is crucial that this data can only be accessible by those authorized to do so. With that in mind, the project's backup and data recovery strategy aims at ensuring that no data loss will occur during the course and after the completion of OPTIMAI, either from human error or hardware failure, as well as inhibit any unauthorized access (see Section 3.6).

All project partners are responsible for processing data within their private servers and will ensure that this data is protected, and any necessary data security controls have been implemented, to minimize the risk of information leak and destruction. This case refers to the data that will be closed and therefore will not be shared and/or re-used within the framework of the project.

The security principles are listed below:

- **Authentication**: All the users wanting to get access to the OPTIMAI datasets should be authenticated. Also, proper means are used to authenticate the servers. An authentication system could be used to handle the authentication of the users and devices during the course of the project.
- **Authorization**: The access to OPTIMAI datasets must be only available to the authenticated and authorized users. These categories and the rights of those users are defined and enforced. The appropriate access control policies and mechanisms (including physical access control) shall be identified for each data set.

- **Accounting:** In OPTIMAI any access and modification to a resource by any user is securely traced/logged in order to prevent users from denying that data files were accessed, altered or deleted, auditing. Other accounting mechanisms shall be implemented.
- **Confidentiality**: The data produced in OPTIMAI should be encrypted during transmission and potentially even in storage.
- **Communication Security**: Access to OPTIMAI portal and repository should be done through encrypted communication channels such as HTTPS and IPsec.
- **Data Integrity**: The data collected during OPTIMAI should be protected from malicious and accidental modifications by any users during their transmission or their storage.

Security will be considered additionally for the purposes of data exchange between partners and sharing before the final data integration/publication.

3.6 Data Storage and Backup

Given the possibility for the OPTIMAI project to manage high volumes of data, a plan for storage and backup of data must be realized, in order to foresee situations of exponential growth of the data volume. Scalability will be guaranteed in order to manage these situations. While for each dataset storage and backup are discussed in detail, OPTIMAI also adheres to common principles to ensure the data is stored correctly backed-up in order to face possible recovery and also that data is secured and governed by access control to ensure consistency and confidentiality.

3.6.1 Storage

This section provides initial considerations for the OPTIMAI data storage in alignment with the FAIR data principles and Open Research Data (ORD) pilot.

Data is to be stored in a secure environment. Sensitive information needs to be stored in the appropriate infrastructure and format, corresponding to the related requirements and specifications of each activity and pilot. Accessibility to the information needs to be maintained and controlled, and the networking configuration should not allow data duplication and circulation. Identification and access to OPTIMAI data complies with national and EU legal requirements and guidance, as well as (in turn) to partner organisations' standards. For this reason, datasets may either be managed by:

- The partner, locally, in its own research data repository, which will be included in a register of research data repositories.
- A trusted and certified repository (i.e., an external infrastructure hosting the data).

Data is to be stored following the guidelines that OPTIMAI Consortium provides during the project development, in order to also be compliant with the ethics and privacy policies. Furthermore, anonymization (and/or pseudonymization) is applied to datasets, if required by their pilots.

Finally, time of storage duration should also be considered for the cases in which data destruction is needed after an establish time period (e.g., after project end).

3.6.2 Backup and Recovery

Backup and recovery strategy will be planned, in order to prevent data loss risks. Confidentiality must be strictly maintained and anonymization will be applied, where required. In addition, partners will not allow to reveal sensitive data about participants to test cases, pilots, trials etc. These same principles will be taken into consideration in the dissemination of data.

3.7 Data archiving and preservation

The DMP will provide guidelines for the short and long term of data archiving and preservation. Data deletion clauses are also considered as part of data archiving policies.

3.7.1 Short Term

The data used and produced during the project development will be updated each time they change in the project lifetime. The new updated dataset will be identified by the number of the previous dataset version plus one, according to the naming convention reported above. For each dataset update, a reference document will also be produced. This document will report the changes of the dataset with respect to previous the version. Also, the reference document will be identified by the number of the previous document version plus one, according to the naming convention previously reported.

Project datasets for training tasks will be generally archived on the OPTIMAI collaborative platform, in order to make them available to all Consortium partners (still through restricted authentication process). In particular cases, especially for confidential or sensitive data, the responsible partners may store these data in their infrastructure.

3.7.2 Long Term

As a project funded by the European Commission, OPTIMAI datasets used in the demonstrator will be maintained for at least five years after project termination. Sensitive data preservation will follow the guidelines that OPTIMAI Consortium will provide during the project development.

In this time, the Consortium will ensure that they remain accessible and usable and destruction of research data will not take place, unless a participant requests it. For cases in which the Consortium will not be able to keep data available for the established time, specific archiving policies for those data will be provided and well documented by the Consortium. The choice of the repository will be carried out considering the present consolidated solutions. The costs required to manage the repository will be divided between the partners and then described and justified.

3.7.3 Data deletion

For both short-term and long-term cases, project data can be deleted only after the related archiving deadlines are reached. Destruction of research data before established deadlines will not take place, unless a participant formally requests it and the Consortium approves the deletion. Additionally, deletion of personal data will be carried out upon triggering of the

situations provided by Art. 17 of the GDPR (Right to erasure).[34] In this case, upon request from the data subject (as defined in the above GDPR article), dataset owner / managing partners (acting as data 'controller') will abide to the GDPR provisions and follow suit.

3.8 Allocation of resources

Regarding the resources related to data management activities, some questions should be considered:

- What additional resources are needed to deliver your plan?
- Is additional specialist expertise (or training for existing staff) required?
- Do you have sufficient storage and equipment or do you need to cost in more?
- Will charges be applied by data repositories?
- Have you costed in time and effort to prepare the data for sharing / preservation?

All resources needed to deliver the plan should be carefully considered and also outlined and justified. Any relevant technical expertise, support and training required and the process to acquire this kind of expertise should be justified in detail as well as any hardware or software or additional storage and backup costs incurred by IT services.

Funding should be included to cover any charges applied by data repositories, for example to handle data of exceptional size or complexity.

The cost in time and effort to prepare data for deposit and the appropriate documentation to enable reuse should be taken into account. The cost data management (in a data repository or in own resources) must be also considered and detailed.

Also remember to cost in time and effort to prepare data for deposit and ensure it is adequately documented to enable reuse. If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data[35].

So, a correct allocation of resources includes:

- estimating the costs for making the data FAIR and describing the methods of covering these costs;
- identifying responsibilities for data management in the project;
- describing costs and potential value of long-term preservation.

3.8.1 Costs for making data FAIR

- Fees associated with the publication of scientific articles containing project's research data in "Gold" Open access journals. The cost sharing, in case of multiple authors, shall be decided among the authors on a case-by-case basis.

- Project Website operation: to be determined.

[34] See https://gdpr-info.eu/art-17-gdpr/

[35] UKDS (UK Data Service) guidance on costing data management. https://www.ukdataservice.ac.uk/manage-data/plan/costing

- Data archiving at ZENODO: the Consortium will evaluate the opportunity that the repository for OPTIMAI open access publications will be based on the free Zenodo service; in this case, no costs are foreseen to make the resulting data and reports generated by OPTIMAI openly available.

In general, the costs for making data FAIR are expected to be covered by the partner's budget allocations.

3.8.2 Responsibilities of partners

Every partner is responsible for the data they produce, i.e., for maintaining and documenting their own datasets as part of the task in which they are working. Where partners, create, generate, consume, process or otherwise use data jointly, partners should identify a 'lead' partner for data management of that specific dataset to ensure a single source of truth and a single version to be archived in order to prevent conflicting copies. Research data will be stored in one of three places (1) project-wide repository for access by all partners but not publicly available; (2) partner internal repository where security restrictions prevent sharing of data between all partners; or (3) dedicated research data repository as appropriate.

Any fee incurred for Open Access through scientific publication of the data will be the responsibility of the data owner (authors) partner(s) in compliance with the CA, Article 8.4.2.1: during the Project and for a period of 1 year after the end of the Project, the dissemination of own Results by one or several Parties including but not restricted to publications and presentations, shall be governed by the procedure of Article 29.1 of the Grant Agreement.

3.8.3 Costs of potential value of long-term preservation

As previously said (cost for making data FAIR), costs of data storage and maintenance are not going to require extra funding once the project ends. From the initial data sources provided by end-users, data may be maintained for the lifetime of the project. This requires databases and access control systems to be in place for the lifetime of the project which is something that will be considered in the architecture and purchasing of OPTIMAI's infrastructure.

As per the value of the data, it is important to take into account that the topics covered by the project respond to a current need of the manufacturing sector and customers' needs. Therefore, data coming out of this project will have a direct impact in the coming years, but might not be of relevance as the challenges are being tackled or replaced by other priorities.

The beneficiaries must give each other access to results needed for exploiting their own results. Requests for access may be made — unless agreed otherwise — up to one year after the project (Grant Agreement Article 31.3).

4 OPTIMAI Datasets Update

This section contains a review of each work package and the associated tasks in turn to identify datasets that are expected to be present within the OPTIMAI project. Each partner was asked to consider their task and any data that is likely to be processed throughout the project and provide the information described in Table *1* using the template reported in Annex III Dataset template. Moreover, each partner started to identify the personal data contained in the datasets filling in a specific template for that available at Annex IV Personal data template. As previously mentioned, data management is an ongoing activity and therefore the information relating to the datasets will be updated over time (particularly where responses are incomplete or only partial at this stage).

4.1 WP1 – Project Management

WP1 relates to the overall management of the project concern the administrative and financial functions. No research datasets are expected to be created during the implementation of the tasks, however, some datasets for management purposes are collected.

Table 4: Overview of tasks and datasets in WP1

Task No	Task Name	Responsible	Related Datasets
T1.1	Project coordination & financial management	CERTH	N/A
T1.2	Scientific coordination and technical management	FORTH	WP1_1 Project Management data
T1.3	Quality assurance and risk management	ENG	WP1_2 OPTIMAI Datasets

Overview	
Dataset ID	WP1_1
Dataset Title	Project Management data
Work Package	WP1
Task / Deliverable	T1.2
Partner(s)	All
Data Type	Project management data
Format	xls/docx

Details	
Description	A set of files pertaining to the technical and scientific management. All the data are stored in the Project nextcloud repository and all the partners have access to it.
Data Size	< 100 MB
Status	Established
Use in OPTIMAI	To monitor and report the technical and scientific developments of the project.
Use beyond OPTIMAI	*N/A*
Open Data	
Is the data open?	No
Explanation	The data concern the internal management process of the project.
Storage location	N/A
Who	The lead beneficiary is FORTH. However, all the partners are eligible to amend and extend the existing data.
Metadata	N/A
How	Through the OPTIMAI nextcloud repository.
Increase data re-use	How and when the data will be made available for re-use
Ethics and Data Protection	
Personal Data	Some of the data include the contact details of the partners' personnel.
Security Requirements	Access to this data has only authorized personnel of each partner. Access is provided after logging in to the nextcloud repository.

Overview	
Dataset ID	WP1_2
Dataset Title	OPTIMAI Datasets
Work Package	WP1
Task / Deliverable	T1.3
Partner(s)	All

Data Type	Project management data
Format	xls
Details	
Description	The dataset contains the list of all the datasets managed in the project. Each partner is providing details on the managed datasets
Data Size	<10 MB
Status	Established
Use in OPTIMAI	The data will be used to monitor and keep track of the datasets managed during the project lifecycle
Use beyond OPTIMAI	Other projects will be able to access the list of datasets of the project and be informed of datasets that they could access and reuse
Open Data	
Is the data open?	YES - but restricted at present
Explanation	The datasets repository is updated and the list is also reported in the Data Management Plan which is a public deliverable
Storage location	OPTIMAI Repository
Who	The responsible partner is ENG but all the partners will be allowed to amend the existing data.
Metadata	Through the project website
How	The dataset will be made available through the Data Management Plan which is a public deliverable and will be published on the website
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	No specific personal data are contained
Security Requirements	At present access is provided after logging in to the nextcloud repository.

4.2 WP2 - User requirements, Technical Specifications and Use case analysis

WP2 concerns the elaboration of the project use cases, the user requirements and the system architecture specifications which are not expected to produce specific datasets themselves but will lay the groundwork for datasets that may be required in other areas of the project.

Table 5: Overview of tasks and datasets in WP2

Task No	Task Name	Responsible	Related Datasets
T2.1	Consolidation of user and ethics and legal requirements	KLEE	N/A
T2.2	State of the art analysis, existing and past research initiatives	UTH	N/A
T2.3	System specifications and architecture	FORTH	N/A
T2.4	Use cases definition	KLEE	N/A

4.3 WP3- Adaptive sensorial network for product-process monitoring and analysis of defects in manufacturing

The objective of WP3 is to design and develop the system hardware infrastructure and software components for monitoring and analysis of defects and faults as well as for security and trust mechanism of data which are provided by the manufacturing domain. A set of files will be collected via suitable sensors in the production line of each pilot, at present these datasets are not related to a specific task and more information will come during the project lifecycle.

Table 6: Overview of tasks and datasets in WP3

Task No	Task Name	Responsible	Related Datasets
--	--	--	WP3_1 Defect_detection_TELEVES WP3_2 Defect_detection_KLEEMANN WP3_3 Defect_detection_MICROSEMI
T3.1	Multisensorial data acquisition and actuation network	EVT	WP3_4, WP3_5 Airquality Sensors Data

T3.2	Middleware for device management and data registration	FINT	WP3_6 Middleware Data, WP3_8 Image data for Kleemann use case 1, WP3_9 Image/3D data for Televes use case 1: Reduce the number of defects
T3.3	Development of OPTIMAI security middleware	FINT	WP3_7 Security middlebox Data
T3.4	On-the-edge processing for acquisition and actuation	ENG	N/A
T3.5	Blockchain framework for traceability and data integrity	CERTH	N/A
T3.6	AI components for quality control towards zero defect manufacturing	UTH	N/A

Overview	
Dataset ID	WP3_1
Dataset Title	Defect_detection_TELEVES
Work Package	WP3
Task / Deliverable	N/A
Partner(s)	TELEVES
Data Type	Derived
Format	Not finalized yet
Details	
Description	A set of files pertaining to the products' inspection to detect defects. Planned to be collected in the production line via suitable sensors (e.g., RGB & Depth camera etc.)
Data Size	Not finalized yet
Status	Planned
Use in OPTIMAI	To minimize defects in the production line (UC1 - Zero defect quality inspection).
Use beyond OPTIMAI	N/A

Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	CERTH
Metadata	N/A
How	N/A
Increase data re-use	*N/A*
Ethics and Data Protection	
Personal Data	The dataset includes personal data of the operators. There is no need or intention to process and store identification data like name, surname, contact details etc. However, considering the motion activities and health –related data (e.g., heart rate), suggested techniques that will be recommended in the context of WP9, will be applied.
Security Requirements	CERTH will collect only the necessary data for research aims, in the context of the project. CERTH will apply technical, regulatory and physical security measures in order to secure the protection of personal data and prevent unauthorized access and fraudulent use. In addition, the security system will be periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP3_2
Dataset Title	Defect_detection_KLEEMANN
Work Package	WP3
Task / Deliverable	N/A
Partner(s)	TELEVES
Data Type	Derived
Format	N/A

Details	
Description	A set of files pertaining to the products' inspection to detect defects. Planned to be collected in the production line via suitable sensors (e.g., noise, vibration, RGB & Depth camera etc.)
Data Size	N/A
Status	Planned
Use in OPTIMAI	To minimize defects in the production line (UC1 - Zero defect quality inspection). Specifically, for KLEEMANN: KPI-PS3.2: Speed-up of the quality inspection process by 30% KPI-PS3.3: Improve final product quality by 5%, as measured by speed, vibrations and noise
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	CERTH
Metadata	N/A
How	N/A
Increase data re-use	*N/A*
Ethics and Data Protection	
Personal Data	The dataset includes personal data of the operators. There is no need or intention to process and store identification data like name, surname, contact details etc. However, considering the motion activities and health –related data (e.g., heart rate), suggested techniques that will be recommended in the context of WP9, will be applied.
Security Requirements	CERTH will collect only the necessary data for research aims, in the context of the project. CERTH will apply technical, regulatory and physical security measures in order to secure the protection of personal data and prevent unauthorized access and fraudulent use. In addition, the security system will be periodically revised in

	order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP3_3
Dataset Title	Defect_detection_MICROSEMI
Work Package	WP3
Task / Deliverable	N/A
Partner(s)	MICROSEMI
Data Type	Derived
Format	N/A
Details	
Description	A set of files pertaining to the products' inspection to detect defects. Planned to be collected in the production line via suitable sensors (e.g., RGB & Depth camera etc.)
Data Size	N/A
Status	Planned
Use in OPTIMAI	To minimize defects in the production line (UC1 - Zero defect quality inspection). Specifically, for MICROSEMI: KPI-PS2.1: Reduce process failure rate from 1% to 0.4%, KPI-PS2.3: Increase production capacity
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	CERTH
Metadata	N/A
How	N/A
Increase data re-use	*N/A*

Ethics and Data Protection	
Personal Data	The dataset includes personal data of the operators. There is no need or intention to process and store identification data like name, surname, contact details etc. However, considering the motion activities and health –related data (e.g., heart rate), suggested techniques that will be recommended in the context of WP9, will be applied.
Security Requirements	CERTH will collect only the necessary data for research aims, in the context of the project. CERTH will apply technical, regulatory and physical security measures in order to secure the protection of personal data and prevent unauthorized access and fraudulent use. In addition, the security system will be periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP3_4
Dataset Title	Not available yet (will be related to dimensional quality control)
Work Package	WP3
Task / Deliverable	T3.1
Partner(s)	UNIMET, MTCL
Data Type	Derived
Format	tbd (.xml, .stl, .qif, .pdf)
Details	
Description	Data as output from 3D scanning processes of manufactured parts (point clouds and annotations that shows the deviation between the manufactured part and the design)
Data Size	N/A
Status	N/A
Use in OPTIMAI	In MTCL use cases
Use beyond OPTIMAI	N/A
Open Data	

Is the data open?	No
Explanation	N/A
Storage location	OPTIMAI repository
Who	Partners involved in the work - end user and technology partners
Metadata	N/A
How	N/A
Increase data re-use	*N/A*
Ethics and Data Protection	
Personal Data	No
Security Requirements	TBD

Overview	
Dataset ID	WP3_5
Dataset Title	Airquality Sensors Data
Work Package	WP3
Task / Deliverable	T3.1
Partner(s)	FINT
Data Type	Derived data
Format	tbd (.json)
Details	
Description	Data as output from the FINoT Indoor Air Quality node equipped with the following sensors Air Temperature, Air Relative Humidity, Total Volatile Organic Compounds (TVOC), and CO2e.
Data Size	TBD
Status	planned
Use in OPTIMAI	The data will be used mainly for quality inspection and early defect detection
Use beyond OPTIMAI	TBD
Open Data	

Is the data open?	No
Explanation	-
Storage location	OPTIMAI repository/middleware
Who	FINT
Metadata	-
How	The data will be available via the middleware API
Increase data re-use	-
Ethics and Data Protection	
Personal Data	No
Security Requirements	TBD

Overview	
Dataset ID	WP3_6
Dataset Title	Middleware Data
Work Package	WP3
Task / Deliverable	T3.2
Partner(s)	FINT, EVT, YBQ, ENG, UNIMET, UPV, KLEE, TVES, MSL
Data Type	Derived data
Format	json
Details	
Description	Data stemming from the deployed sensors so as long as the sensor data do not contain any personal data the middleware will not process/store any type of personal data
Data Size	tbd
Status	planned
Use in OPTIMAI	The data will be used mainly for quality inspection and early defect detection
Use beyond OPTIMAI	tbd
Open Data	

Is the data open?	No
Explanation	N/A
Storage location	OPTIMAI repository/middleware
Who	FINT and Partners involved in the work
Metadata	-
How	The data will be available via the middleware API
Increase data re-use	-
Ethics and Data Protection	
Personal Data	No
Security Requirements	The middleware can only be accessed with trusted credentials complying to the authorisation level of each authenticated requesting entity. Furthermore, the interactions with the middleware are allowed only via TLS based encrypted communication channels

Overview	
Dataset ID	WP3_7
Dataset Title	Security middlebox Data
Work Package	WP3
Task / Deliverable	T3.3
Partner(s)	FINT, CERTH, EVT, YBQ, ENG, UPV, TRI
Data Type	Derived data
Format	json
Details	
Description	The ingress and egress network traffic will be analysed for detecting anomalies that indicate cybersecurity threats
Data Size	tbd
Status	planned
Use in OPTIMAI	The security middlebox needs to be able to detect and mitigate cyber threats, in this context and depending on the deployed

	security solutions the analysis of data like IP addresses, network traffic, etc is a requirement.
Use beyond OPTIMAI	tbd
Open Data	
Is the data open?	No
Explanation	-
Storage location	Security Middlebox, DSS or other analysis platform (if any, tbd)
Who	FINT Partners involved in the work
Metadata	-
How	The data will be available via the middlebox API and/or dashboard
Increase data re-use	-
Ethics and Data Protection	
Personal Data	No
Security Requirements	The security middlebox can only be accessed with trusted credentials complying to the authorisation level of each authenticated requesting entity. Furtermore the interactions with the middlebox ar allowed only via SSH and HTTPS based encrypted communication channels

Overview	
Dataset ID	WP3_8
Dataset Title	Image data for Kleemann use case 1
Work Package	WP3 (also related to WP7)
Task / Deliverable	T3.2 (also related to T7.3)
Partner(s)	EVT, KLEEMANN
Data Type	Image data used for prototyping a sensor
Format	images, text
Details	
Description	Images and labeling information at the very beginning stage of the sensor definition period.

Data Size	Not yet known
Status	Data bases of images from the process.
Use in OPTIMAI	Used for manual /semi-automatic or automatic configuration and training of the sensors.
Use beyond OPTIMAI	The data is rleated to the specific pilot use cases and belongs to the pilot partners. EVT won't share the data with external researchers.
Open Data	
Is the data open?	No
Explanation	Data comes from the OPTIMAI use cases
Storage location	At this stage of the project, the data will be manually provided by project partners. EVT will respect the project partner's conventions and requirements. For example, if the project partner allows access only via sfpt.
Who	EVT and the according project partner will find ways to share the data. Preparation is still ongoing
Metadata	It is planned that EVT will write a list about data bases that they are using inside the "sensor_installation" folder inside the OPTIMAI file server.
How	Image data can be accessed with any image processing program, 3D data can be accessed with tools of EVT, but may also be converted into other formats.
Increase data re-use	Not yet planned
Ethics and Data Protection	
Personal Data	No
Security Requirements	No

Overview	
Dataset ID	WP3_9
Dataset Title	Image/3D data for Televes use case 1: Reduce the number of defects
Work Package	WP3 (also related to WP7)

Task / Deliverable	T3.2 (also related to T7.4)
Partner(s)	EVT, TELVES
Data Type	Image data used for prototyping a sensor
Format	images, text
Details	
Description	Images and labelling information at the very beginning stage of the sensor definition period.
Data Size	Not yet known
Status	Data bases of images from the process.
Use in OPTIMAI	Used for manual /semi-automatic or automatic configuration and training of the sensors.
Use beyond OPTIMAI	The data is related to the specific pilot use cases and belongs to the pilot partners. EVT won't share the data with external researchers.
Open Data	
Is the data open?	No
Explanation	Data comes from the OPTIMAI use cases
Storage location	At this stage of the project, the data will be manually provided by project partners. EVT will respect the project partner's conventions and requirements. For example, if the project partner allows access only via sfpt.
Who	Telves is sending parts to EVT, EVT will do a preliminary evaluation.
Metadata	It is planned that EVT will write a list about data bases that they are using inside the "sensor_installation" folder inside the OPTIMAI file server.
How	Image data can be accessed with any image processing program, 3D data can be accessed with tools of EVT, but may also be converted into other formats.
Increase data re-use	Not yet planned
Ethics and Data Protection	
Personal Data	No

Security Requirements	No

4.4 WP4 - Production virtualization

The goal of WP4 is to develop software components for the production optimization process. A dataset containing virtual models, already used in other projects, will be the starting point to create simulation models and to virtualize production scenarios. The datasets identified so far (M6) are described in the following tables.

Table 7: Overview of tasks and datasets in WP4

Task No	Task Name	Responsible	Related Datasets
T4.1	Digital twinning of manufacturing processes	VIS	N/A
T4.2	Sensors modelling and virtualization	FINT	WP4_2 Virtualised Airquality Sensors Data
T4.3	Prediction of upcoming defects towards zero downtime manufacturing	CERTH	N/A
T4.4	Simulation engine for production planning	VIS	WP4_1 eCat

Overview	
Dataset ID	WP4_1
Dataset Title	eCat
Work Package	WP4
Task / Deliverable	T4.4
Partner(s)	VIS
Data Type	• Primary data collected by partner • Secondary data (not publicly available)
Format	.vcmx, .vcm
Details	
Description	The virtual components have been created and maintained by VIS and contains the following datasets: Geometries (CAD datasets), Behaviours and Parameters (design and operation datasets)

Data Size	~2GB
Status	Established
Use in OPTIMAI	The available virtual models will be the starting point to create simulation models and to virtualize production scenarios.
Use beyond OPTIMAI	Virtual models are currently used within different research projects. The utilization of the models required a valid license of Visual Components 4.0, and the acceptance of the EULA.
Open Data	
Is the data open?	No, available only with a valid license of Visual Components 4.0 and acceptance of the EULA
Explanation	
Storage location	
Who	Visual Components Oy
Metadata	Data only accessible through Visual Components 4.0 GUI
How	Data only accessible through Visual Components 4.0 GUI software
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	No personal data is included
Security Requirements	eCat is located in a secure server with only read functionality. Maintenance of the virtual components is done by Visual Components Oy.
Comments	The public eCat is only available for the users with a valid license for Visual Components 4.0 which have accepted the EULA before installing.

Overview	
Dataset ID	WP4_2
Dataset Title	Virtualised Airquality Sensors Data
Work Package	WP4
Task / Deliverable	T4.2

Partner(s)	FINT
Data Type	Synthetic/generated
Format	tbd (.json)
Details	
Description	Generate predictions regarding the physical readings of the FINoT Indoor Air Quality sensors Air Temperature, Air Relative Humidity, Total Volatile Organic Compounds (TVOC), and CO2e.
Data Size	tbd
Status	planned
Use in OPTIMAI	The difference between the expected/estimated measurements and the real value will be used as indicative of the physical sensor health and to also predict in nearer time any possible failures in the process.
Use beyond OPTIMAI	tbd
Open Data	
Is the data open?	No
Explanation	-
Storage location	OPTIMAI repository/middleware
Who	FINT
Metadata	-
How	The data will be available via the middleware API
Increase data re-use	-
Ethics and Data Protection	
Personal Data	No
Security Requirements	tbd

4.5 WP5 - Novel techniques and systems for fast production (re)-configuration and planning

WP5 focuses on deploying novel tools and methodologies to be used for the quality monitoring process. To support AR functionalities a set of files pertaining the operators' activity during the execution of a task at the production line of each pilot will be collected. These datasets won't contain any contact details about the operators but information concerning the motion activities and health related data (e.g. heart rate), therefore the principles of the ethics and legal framework will be applied. Moreover technical, regulatory and physical security measures in order to secure the protection of personal data and prevent unauthorized access and fraudulent use will be adopted. The partner YBQ will store in the internal cloud server a dataset containing contact details of the partners for communication purposes.

Table 8: Overview of tasks and datasets in WP5

Task No	Task Name	Responsible	Related Datasets
--	--	--	WP5_1 activity_recognition_TELEVES WP5_2 activity_recognition_KLEEMANN WP5_3 activity_recognition_MICROSEMI
T5.1	Perception techniques for operator-machine interaction	CERTH	N/A
T5.2	Augmented reality interface and visual analytics	FORTH	N/A
T5.3	Wearable devices for real-time assistance on the production line	YBQ	WP5_4 Contact Details
T5.4	On-the-fly production (re)-configuration techniques	FORTH	N/A

Overview	
Dataset ID	WP5_1
Dataset Title	activity_recognition_TELEVES
Work Package	WP5
Task / Deliverable	

Partner(s)	TELEVES
Data Type	Derived, Synthetic data
Format	N/A
Details	
Description	A set of files pertaining the operators' activity during the execution of a task at the production line. Data planned to be collected via suitable sensors (e.g., RGB & Depth camera), installed on the operators' AR glasses, or/and at specific positions on the production line. Semantic segmentation will be performed in real-time in order to separate a particular object of interest (produced parts or production machines) from its background and subsequently estimate its pose. The objects' relative position with respect to a human operator will be recognised. A semantic fusion of quality control information (e.g., identified defects - defect_detection_TELEVES dataset, metrological measurements, operator's heart rate etc.) will be executed and propagated to the foreseen AR environment.
Data Size	N/A
Status	Planned
Use in OPTIMAI	To support AR functionalities (UC 2 - Production line setup-calibration). Specifically, for TELEVES speed up operator's decision process and interaction with machine etc.
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	CERTH
Metadata	N/A
How	N/A
Increase data re-use	N/A
Ethics and Data Protection	

Personal Data	The dataset includes personal data of the operators. There is no need or intention to process and store identification data like name, surname, contact details etc. However, considering the motion activities and health –related data (e.g., heart rate), suggested techniques that will be recommended in the context of WP9, will be applied.
Security Requirements	CERTH will collect only the necessary data for research aims, in the context of the project. CERTH will apply technical, regulatory and physical security measures in order to secure the protection of personal data and prevent unauthorized access and fraudulent use. In addition, the security system will be periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP5_2
Dataset Title	activity_recognition_KLEEMANN
Work Package	WP5
Task / Deliverable	
Partner(s)	KLEEMANN
Data Type	Derived, Synthetic data
Format	N/A
Details	
Description	A set of files pertaining the operators' activity during the execution of a task at the production line. Data planned to be collected via suitable sensors (e.g., RGB & Depth camera), installed on the operators' AR glasses, or/and at specific positions on the production line. Semantic segmentation will be performed in real-time in order to separate a particular object of interest (produced parts or production machines) from its background and subsequently estimate its pose. The objects' relative position with respect to a human operator will be recognised. A semantic fusion of quality control information (e.g., identified defects - defect_detection_KLEEMANN dataset, metrological measurements, operator's heart rate etc.) will be executed and propagated to the foreseen AR environment.

Data Size	N/A
Status	Planned
Use in OPTIMAI	To support AR functionalities (UC 2 - Production line setup-calibration). Specifically, for KLEEMANN: KPI-PS3.1: Speed up calibration of the valve block by 40%
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	CERTH
Metadata	N/A
How	N/A
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	The dataset includes personal data of the operators. There is no need or intention to process and store identification data like name, surname, contact details etc. However, considering the motion activities and health –related data (e.g., heart rate), suggested techniques that will be recommended in the context of WP9, will be applied.
Security Requirements	CERTH will collect only the necessary data for research aims, in the context of the project. CERTH will apply technical, regulatory and physical security measures in order to secure the protection of personal data and prevent unauthorized access and fraudulent use. In addition, the security system will be periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP5_3
Dataset Title	activity_recognition_MICROSEMI

Work Package	WP5
Task / Deliverable	
Partner(s)	MICROSEMI
Data Type	Derived, Synthetic data
Format	N/A
Details	
Description	A set of files pertaining the operators' activity during the execution of a task at the production line. Data planned to be collected via suitable sensors (e.g., RGB & Depth camera), installed on the operators' AR glasses, or/and at specific positions on the production line. Semantic segmentation will be performed in real-time in order to separate a particular object of interest (produced parts or production machines) from its background and subsequently estimate its pose. The objects' relative position with respect to a human operator will be recognised. A semantic fusion of quality control information (e.g., identified defects - defect_detection_MICROSEMI dataset, metrological measurements, operator's heart rate etc.) will be executed and propagated to the foreseen AR environment.
Data Size	N/A
Status	Planned
Use in OPTIMAI	To support AR functionalities (UC 2 - Production line setup-calibration). Specifically, for MICROSEMI: KPI-PS2.2: Reduce machine setup time by 70%.
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	CERTH
Metadata	N/A
How	N/A
Increase data re-use	N/A

Ethics and Data Protection	
Personal Data	The dataset includes personal data of the operators. There is no need or intention to process and store identification data like name, surname, contact details etc. However, considering the motion activities and health –related data (e.g., heart rate), suggested techniques that will be recommended in the context of WP9, will be applied.
Security Requirements	CERTH will collect only the necessary data for research aims, in the context of the project. CERTH will apply technical, regulatory and physical security measures in order to secure the protection of personal data and prevent unauthorized access and fraudulent use. In addition, the security system will be periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP5_4
Dataset Title	Contact Details
Work Package	WP5
Task / Deliverable	T5.3
Partner(s)	YBQ
Data Type	Project management data
Format	xls
Details	
Description	The dataset will contain partners data for project communication. Data are collected with direct email communication
Data Size	
Status	Established
Use in OPTIMAI	For the project communication of results
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No

Explanation	N/A
Storage location	Company internal cloud server
Who	YBQ
Metadata	N/A
How	Spreadsheet software
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	The dataset contains the email of the project partners
Security Requirements	The internal cloud is private and can only be accessed with trusted credentials

4.6 WP6 - Decision support and system integration

WP6 will create the OPTIMAI Decision Support System for early notifications regarding defects as well as will provide the integrated solution and prepare the ground for realistic system testing, user engagement and evaluation of use-cases. At the current stage no datasets are established, more details will be provided in the next release of the document.

Table 9: Overview of tasks and datasets in WP6

Task No	Task Name	Responsible	Related Datasets
T6.1	Decision support and early notification framework	CERTH	N/A
T6.2	Intelligent Marketplace for AI sharing and scrap re-use	FINT	WP6_1 Marketplace data
T6.3	Testing and incremental integration of components	ENG	N/A
T6.4	System validation	FORTH	N/A

Overview	
Dataset ID	WP6_1
Dataset Title	Marketplace data

Work Package	WP6
Task / Deliverable	T6.2
Partner(s)	FINT, CERTH, FORTH
Data Type	Derived data
Format	tbd (.json)
Details	
Description	Data describing the offered AI algorithms developed within OPTIMAI and b) the defective parts of an Industry
Data Size	tbd
Status	planned
Use in OPTIMAI	The offered AI algorithms will be offered to third parties that would like to increase their production quality by minimised scrap using OPTIMAI Identification (of potential defective parts) and prediction of defects and malfunctions algorithms whereas the offered defective parts could be used from an industry that for them are not scrap but useful (prior to the recycling process or after it).
Use beyond OPTIMAI	tbd
Open Data	
Is the data open?	No
Explanation	-
Storage location	OPTIMAI marketplace
Who	FINT and Partners involved in the work
Metadata	-
How	The data will be available via the marketplace API and/or dashboard
Increase data re-use	-
Ethics and Data Protection	
Personal Data	Yes, the dataset could include personal data of the AI algorithms owners, the contact points of the companies offering the defective parts and of the employees of the companies that may

	want to purchase the defective parts or the offered AI algorithms. In this context, the suggested techniques that will be recommended in the context of WP9, will be applied.
Security Requirements	The marketplace can only be accessed with trusted credentials complying to the authorisation level of each authenticated requesting entity. Furthermore the interactions with the marketplace allowed only via TLS based encrypted communication channels.

4.7 WP7 - Industrial demonstrator

WP7 is dedicated to the preparation of training material for OPTIMAI end-users, establishing the regulatory framework for the pilot deployment, planning and executing OPTIMAI's industrial demonstrators. Specific datasets for the collection and analysis of information from the plant will be identified in the following months, a first example has been provided hereafter.

Table 10: Overview of tasks and datasets in WP7

Task No	Task Name	Responsible	Related Datasets
T7.1	OPTIMAI Training	CARR	N/A
T7.2	Ethics recommendations & regulatory framework	TRI	N/A
T7.3	Demo in KLEE	KLEE	WP7_1 Sensor Datasets KLEE
T7.4	Demo in TELEVES	TVES	N/A
T7.5	Demo in MICROSEMI	MTCL	N/A

Overview	
Dataset ID	WP7_1
Dataset Title	Sensor Datasets KLEE
Work Package	WP7
Task / Deliverable	T7.3
Partner(s)	KLEE, CERTH, FINT, FORTH, EVT, VIS, YBQ, UTH, ENG, UPV, CARR, UAB

Data Type	Derived data (e.g., output from processing by OPTIMAI module)
Format	N/A
Details	
Description	N/A
Data Size	N/A
Status	Planned
Use in OPTIMAI	The data will be used mainly for quality inspection and early defect detection. The dataset will also assist the (re)configuration of the hydraulic units' parameters and the production planning
Use beyond OPTIMAI	The dataset will be useful for other researchers and also end-users, since it will combine digital twins with AI models. Based on the above combination, the data will provide real-world use cases that will integrate and validate the OPTIMAI solutions and allow for end-user feedback.
Open Data	
Is the data open?	No
Explanation	
Storage location	
Who	
Metadata	
How	Via OPTIMAI platform
Increase data re-use	*N/A*
Ethics and Data Protection	
Personal Data	If personal data are to be used, a consent form will be given.
Security Requirements	OPTIMAI security measures

4.8 WP8 - Dissemination, commercialization and exploitation strategies

WP8 is raising public awareness of project achievements among the key user groups and stakeholders and is developing the exploitation plan and the roadmap to the market.

A dataset containing the newsletter recipients is foreseen and it will be used for dissemination purposes. One internal dataset collects the exploitation contacts will be used for managing and communication purposes. Moreover, a dataset for key exploitable results is planned and will be established in the next month of the project.

Table 11: Overview of tasks and datasets in WP8

Task No	Task Name	Responsible	Related Datasets
T8.1	High impact dissemination & communication activities	CARR	WP8_1 OPTIMAI Newsletter Recipients
T8.2	Clustering, networking & Knowledge transfer activities	TRI	N/A
T8.3	Standardization activities	KLEE	N/A
T8.4	Market analysis and segmentation	TVES	N/A
T8.5	Knowledge management and IPR protection	UNIMET	N/A
T8.6	Exploitation plan and roadmap to market	ENG	WP8_3 Key Exploitable Results Dataset

Overview	
Dataset ID	WP8_1
Dataset Title	OPTIMAI Newsletter Recipients
Work Package	WP8
Task / Deliverable	T8.1
Partner(s)	CARR
Data Type	Primary data collected by CARR in OPTIMAI
Format	CSV
Details	
Description	Dataset will consist of names and email addresses of those who voluntarily sign-up to the OPTIMAI Newsletter via the OPTIMAI website
Data Size	N/A
Status	Planned

Use in OPTIMAI	Solely to distribute the OPTIMAI Newsletter
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	The names and personal information will be stored on the servers of the Mailchimp newsletter platform which will be used by the OPTIMAI project
Who	N/A
Metadata	N/A
How	N/A
Increase data re-use	*N/A*
Ethics and Data Protection	
Personal Data	*Yes, informed consent will be given via the sign-up sheet for the OPTIMAI Newsletter*
Security Requirements	Access to the dataset will be limited to two people within the Carr Communications team, Communication and Dissemination Manager and Executive. All personal data will be password protected and will not be shared beyond the two mentioned individuals.

Overview	
Dataset ID	WP8_3
Dataset Title	Key Exploitable Results Dataset
Work Package	WP8
Task / Deliverable	T8.6
Partner(s)	All
Data Type	Primary data collected by ENG during the OPTIMAI project activities
Format	xls/docx

Details	
Description	A list of key exploitable results of the OPTIMAI project. All the data are stored in the Project nextcloud repository and all the partners have access to it.
Data Size	< 10 MB
Status	Planned
Use in OPTIMAI	To identify and prioritise the main interesting results of the project for exploitation purposes
Use beyond OPTIMAI	*N/A*
Open Data	
Is the data open?	Yes, partially
Explanation	The dataset will be used internally but the relevant information will be inserted in the corresponding deliverables.
Storage location	OPTIMAI repository
Who	The lead beneficiary is ENG. However, all the partners are eligible to amend and extend the existing data.
Metadata	N/A
How	Through the OPTIMAI nextcloud repository.
Increase data re-use	The relevant information will be included in the public deliverable which will be made available on the project website.
Ethics and Data Protection	
Personal Data	No personal details are foreseen at present
Security Requirements	Access is provided after logging in to the nextcloud repository.

4.9 WP9 - Legal and Ethical framework

The objective of WP9 is to set up a monitoring and enforcement strategy to ensure that the outcomes of the project comply with the legal and ethical requirements. At present no dataset are foreseen.

Table 12: Overview of tasks and datasets in WP9

Task No	Task Name	Responsible	Related Datasets
T9.1	Identification of the ethical and legal framework	UAB	N/A
T9.2	Identification and analysis of the ethical, legal and societal risks	TRI	N/A
T9.3	Definition and implementation of the OPTIMAI Regulatory Model	UAB	N/A
T9.4	Ethical and Legal monitoring	TRI	N/A

5 Conclusions

This deliverable has set out the principles for data management within OPTIMAI. The OPTIMAI DMP aims at safeguarding the sound management of the data produced/used during the project's activities across their entire lifecycle. The data within OPTIMAI is expected to adhere to the FAIR principles and the ways that data will be made findable, accessible, interoperable, and reusable have been detailed within this report. This document describes all the underlying processes of the OPTIMAI data management, collection and generation, in accordance with the GDPR guidelines and the FAIR principles. In addition, this document sheds light the data foreseen to be collected/generated under the project activities, methodologies to make dataset FAIR, the allocation of resources and the data security. Moreover, this document deals with procedures for storage and preservation of dataset and it describes the ethical aspects to be addressed during the project. Each WP has been analysed for data processing activities and relevant datasets have been identified. The ethical and legal aspects are defined in conjunction with the activities of WP9 where the ethical and legal framework of OPTIMAI is established.

It is recognised that the data management within the project should be an ongoing process that should be reviewed and updated throughout the project's timeline, the next release of the Data Management Plan will be delivered at M24 (D1.4 – Data Management Plan -3rd version).

References

[1] European Commission. (2016). H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. Available at: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[2] Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1-9

[3] GO FAIR. (2019). FAIR Principles. from GO FAIR: https://www.go-fair.org/fair-principles/ (Last Accessed June 2021)

[4] 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 at: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[5] ALLEA. (2017). The European Code of Conduct for Research Integrity – ALLEA, All European Academies, Berlin 2017. Available at https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf (Last Accessed June 2021).

[6] Council of Europe/European Court of Human Rights. (2021). Guide to the Case-Law of the of the European Court of Human Rights: Data Protection. Available at: https://www.echr.coe.int/Documents/Guide_Data_protection_ENG.pdf

[7] European Data Protection Board. (2020). Guidelines 07/2020 on the concepts of controller and processor in the GDPR. Available at: https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_202007_controllerprocessor_en.pdf

Annex I Informed Consent forms template to be used at internal workshops and similar events

This informed consent form template serves as a guide for OPTIMAI Partners to be used before conducting internal events with the OPTIMAI Consortium. This template must be adapted for each event. WP9 will assist OPTIMAI Partners in adapting this form in accordance with the provisions laid down in the GDPR.

Regarding piloting activities, given the ethical and legal risks that human participation and data processing activities could pose, a specific informed consent form will be designed to ensure genuine voluntary and informed consent. Such informed consent form will be translated into the languages of research participants.

OPTIMAI INFORMED CONSENT FORM TEMPLATE FOR INTERNAL EVENTS

PART I: Informed Consent Information Sheet

Purpose of the *[insert type of research activity, e.g., workshop]*

The *[insert type of research activity, e.g., workshop]* will take place on *[insert date*] *[select as appropriate: via Zoom/Webex/Teams/Gotomeeting …. / in [name of the city]*]. The aim of this *[insert type of research activity],* is *[insert a detailed explanation including: i) the research activity and its aim; ii) types of data that will be collected and the purpose(s) of data collection].*

Personal data from the participants will be kept under the responsibility of the controller, *[insert the name of the partner in charge of the research activity]*, and will not be shared or transferred to partners others than those directly involved in the development of the research activity.

[*Explain whether anonymisation or pseudonimisation techniques will be applied to the collected personal data and provide a detailed explanation of the technique(s) used*]

[*select as appropriate: Anonymised/pseudonymised/personal*] data will be gathered on *[select as appropriate: paper/computer files* stored in our offices at *[insert the name of the partner in charge of the research activity]* premises / the cloud server [*specify which one and whether it entails international data transfers*]*]*. It will be stored securely and confidentially until the project activities have been completed (December 2023).

Rights, voluntary participation and right to withdraw

You can exercise the following rights at any time by contacting the partner responsible for conducting this research activity ([*insert name of the partner*]). To this end, contact *[Name and surname] via [email].*

• The **right of access** to the data we have collected from you: You have the right to be informed as to whether or not your personal data is being processed.

• The **right to rectification** of the data you have provided us: You have the right to request rectification of any mistakes in your personal data.

• The **right to erasure**: You have the right to request to have your personal data erased/deleted.

• The **right to the restriction of processing your data**: You have the right to limit the processing of your personal data.

• The **right to object**: You have the right to object to the data processing in order to stop the processing of your personal data.

Your participation in the *[insert type of research activity]* is completely voluntary. You are free to withdraw from the *[insert type of research activity]* at any time, without giving a reason for your withdrawal and without any consequences.

PART II: Informed Consent Form to participate in OPTIMAI [insert type of research activity]

I hereby declare that:

1. I have carefully read and understood the information provided regarding the *[insert type of research activity]*.

2. I am fully aware of all my rights and, especially, of my right to withdraw consent at any time without consequences by contacting the partner responsible for conducting this research activity. To this end, contact *[Name and surname]* via *[email].*

Additionally, complete this consent form by ticking the boxes below. By ticking "yes", you are consenting to the corresponding data processing activity. By ticking "no", you do not consent to such data processing activity.

[Adapt the following form to the research activity, the types of data to be collected and the purposes (if more than one)]

	YES	NO
1. I consent to the video recording of the *[insert type of research activity]*.	☐	☐
2. I consent to the audio recording of the *[insert type of research activity].*	☐	☐
3. I consent to have my picture taken during the *[insert type of research activity]*	☐	☐
4. *[Add as necessary]*	☐	☐

By signing below, I acknowledge that I will participate in the *[insert type of research activity]* voluntarily.

Name and surname of the participant: .....

Date: .....

Participant's signature:.....

Annex II Data Minimisation Explainer

Data minimisation is required under Article 5(1)(c) of the GDPR. It states:

> personal data shall be: …(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation').

The GDPR does not provide a wider explanation of this paragraph, or these terms, specifically. The data minimisation principle requires you to identify the minimum amount of personal data which is required to fulfil the tasks you will be carrying out in OPTIMAI and hold no more data than that.

Personal Data

Personal data is defined in Art.4(1) of the GDPR as:

> any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Therefore, personal data is that which concerns an individual and can be used to identify them. Where your actions in OPTIMAI process personal data, you must engage in data minimisation.

Pseudonymisation of data may help to reduce privacy risks, but it is still personal data as it may be possible to identify the individual in question. Anonymised data is not personal data and is not subject to the GDPR. Therefore, anonymised data does not need to go through a process of data minimisation. You should be able to demonstrate how your anonymisation techniques work so that the data in question is no longer personal.

This document will now explain each of the three aspects of the data minimisation principle.

Adequate

Personal data is seen to be '*adequate*' where it is sufficient to fulfil your stated purpose; i.e. the amount of data you process is enough to carry out your tasks.[36] For example, in order to obtain end-user requirements it may be necessary to record video of employees engaging in professional activities at work.

[36] Information Commissioner's Office, Principle (c): Data Minimisation, 2019, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/ (Hereafter: ICO).

Relevant

Personal data is seen as '*relevant*' where it has a rational link to your purpose; i.e. the data you process is clearly connected to the task you are performing. For example, video recordings of employees engaged in professional activities may be required for end-user requirement collection, but not video-recordings of employees in other contexts such as their lunch break.

Limited

Personal data is seen a '*limited*' where it is only that which is necessary to fulfil your purpose; i.e. you hold no more data than you need to complete your task. For example, video recordings of employees engaged in their work practices should be limited to as few instances as necessary and not expansive and ongoing.

Whether the data you have is adequate, relevant, and limited must be judged on a case-by-case basis (potentially down to the individual level) and will depend upon the purpose for which you are collecting and processing the data. You must, therefore, be clear as to why you need a particular type of data before you assess it.

If you are collecting or processing special category data, such as that related to trade union membership, you will need to take special care to ensure that the data is thoroughly assessed in terms of the data minimisation principle.

You will also need to take particular care if individuals object to their data being collected or processed, if they request that their data is made accurate, or they request that any unnecessary data is deleted.

Implementation

In terms of implementing the data minimisation principle in the project, you first need to carefully consider what data you really need. This is not intended to constrain your actions when working on a task, but to enable you to act lawfully. Thinking about your data needs can result in designs "*that* [require] *significantly less data, or may require no personal data at all*."[37]

Generally, minimisation of processing personal data can be achieved by either collecting data on less people or collecting less data about relevant people.[38] For example, if you are engaging in surveys and interview, you could consider if you need to hold and process the names of participants in these activities.

[37] Hoepman Jaap-Henk, Privacy Design Strategies, 2019 [p.5]
[38] Hoepman Jaap-Henk, Privacy Design Strategies, 2019 [p.5]

In order to engage in data minimisation, you should:

1. Determine what personal data is needed to carry out your task and process only that. You may include data that is foreseeably relevant to your purpose, but not data that has a low chance of being useful in the future.[39] You should be conservative in terms of determining what data may be necessary,[40] and should be able to justify your choices about the data you plan to use.[41]
2. Determine how long you expect you will need the personal data for prior to collecting or processing it. Protocols for deleting that data should be instigated so that the data is destroyed at the agreed time. These time frames should be reviewed periodically. They may be shortened if the usefulness of the data is expected to be completed sooner than anticipated. They should only be extended where retaining the data is necessary for completing the task. Where the data is part of a larger dataset, it may be changed to be unspecified values rather than deleted (but the entire dataset should be eventually planned for deletion).[42]
3. Exclude data that is irrelevant. Do not plan to collect or process personal data that you do not need for carrying out your task. If this data is provided to you in error, you should destroy it immediately.[43]
4. Periodically review the data you hold, and your processing methods to ensure that they are adequate, relevant, and limited. Reviews could occur following a passage of time, for example every year, and/or after substantive events, such as after data sets have been expanded. Where data are found to be unnecessary after review, they should be destroyed.
5. Remove all personal data as soon as it is no longer useful. This should include back-ups, metadata, and traces of the data. Ensure that the data is not recoverable.[44] Note that destruction of data relates to its presence on the physical storage layer, and not just removing the data from software applications (data stripping).[45]

Note that the use of techniques such as data mining, deep learning, or big data may create new personal data, or new insights into personal data. If you are using such techniques, you should be careful to ensure that only that data which is necessary is included when using these methods, and that only those insights generated from these methods that are necessary to retain are used.[46]

[39] ICO
[40] Hoepman Jaap-Henk, Privacy Design Strategies, 2019 [p.5]
[41] ICO
[42] Hoepman Jaap-Henk, Privacy Design Strategies, 2019 [p.5]
[43] Hoepman Jaap-Henk, Privacy Design Strategies, 2019 [p.5]
[44] Hoepman Jaap-Henk, Privacy Design Strategies, 2019 [p.5-6].
[45] Hoepman Jaap-Henk, Privacy Design Strategies, 2019 [p.7].
[46] Hoepman Jaap-Henk, Privacy Design Strategies, 2019 [p.6].

Annex III Dataset template

Overview	
Dataset ID	
Dataset Title	
Work Package	
Task / Deliverable	*(if applicable)*
Partner(s)	
Data Type	• *Project management data* • *Primary data collected by partner in OPTIMAI* • *Secondary data (not publicly available)* • *Derived data (e.g., output from processing by OPTIMAI module)* • *Publicly available dataset (e.g., training / benchmark data)* • *Synthetic / generated data*
Format	*xls/csv/etc*
Details	
Description	*Brief description of the dataset and how it was collected*
Data Size	
Status	*Specify if it is already in place, established, or planned*
Use in OPTIMAI	*How the data is/will be used in OPTIMAI*
Use beyond OPTIMAI	*How the data could be useful to other researchers beyond OPTIMAI*
Open Data	
Is the data open?	*Yes – Public; Yes – but restricted access, No*
Explanation	*Justification of open access decision*
Storage location	*Where will it be stored; (if open, specify repository)*
Who	*Who is responsible for storing the data*
Metadata	*What metadata has been created and how is this managed*
How	*How can the data be accessed? (software, techniques)*
Increase data re-use	*How and when will the data be made available for re-use?*

Ethics and Data Protection	
Personal Data	*Does the dataset contain personal data? Was informed consent given for use/reuse?*
Security Requirements	*Indicate if there are specific security measures (both technical and organizational) to be considered regarding the dataset*

Annex IV Personal data template

Workpackage/Task	
Dataset ID	
Types of personal data to be processed	*You can indicate the general type of data, e.g. business-related contacts, contact information related to dissemination, personal information from workshops, questionnaires, does the information falls under the category of special categories of data, etc.*
Data Source	*You can note here whether the data will come directly from the data subject or you will receive it from a data base, whether the data will be originated within OPTIMAI or priorly.*
Purpose	*Why do you need this specific type of information*
Legal basis	*Under which legal ground as outlined by Art. 6/ Art. 9 GDPR do you process the personal information*
Data Minimisation	*How do you guarantee that you're not going the collect more information than you need?*
Main controller(s)	*Who determines why and how the data is processed?*
Processors involved	*Are you involved in this activity on your own or you're working with other partners as well?*
Joint controllership	*Do you determine the purpose and the means for data processing with any other partner? If yes, with whom?*
Data Recipients	*Is this information going to be shared? With whom?*
Applicable safeguards	*What technical and organisational measures are in place to ensure a high-level of protection to personal data?*