D1.5

Data Management Plan – 4th version

30th June 2024

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	D1.5: Data Management Plan – 4th version		
Work Package	WP1 – Project Management		
Nature	ORDP: Open Research Data Pilot	**Dissemination Level**	Public
Lead Beneficiary	ENG		
Authors	Cinzia Rubattino (ENG) Elisa Rossi (ENG)		
Co-authors	Agata Gurzawska (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	10/06/2024	Updated ToC	Cinzia Rubattino (ENG) Elisa Rossi (ENG)
V0.2	28/06/2024	All sections updated	Cinzia Rubattino (ENG) Elisa Rossi (ENG) Jorge González-Conejero (UAB) Anaïs Resseguier (TRI)
V0.3	02/07/2024	Dataset Setcion updated	All partners
V0.4	08/07/2024	Internal reviewed version	Jorge González-Conejero (UAB) Sebastian Pantoja (TELEVES)
V1.0	09/07/2024	Final version for submission	Cinzia Rubattino (ENG) Elisa Rossi (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 fourth and final version of the Data Management Plan (DMP) for the OPTIMAI project, labelled as D1.5 - Data Management Plan – 4th version, and it is being submitted in month 42 of the project. The purpose of this document is to streamline the data management process for all data collected, stored, processed, or generated by the project, with the aim of maximizing its accessibility. This is in line with the H2020 Pilot on Open Research Data (ORDP) in which the project is involved.

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 is in full compliance with the EU legislative and regulatory framework for ethics and data protection. While the main regulations and basic concepts of the EU legal framework were summarized in previous versions of this deliverable, the present document only highlights the major concepts. As well, the Data Management Plan Methodology defined in previous versions is here reported to have a self-reading deliverable: 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 are briefly reported, while the complete legal and ethical framework of OPTIMAI are reported in WP9 deliverables.

The final part of the report then reviews each work package and task for any related data management requirement identified. A total of 41 datasets have been identified, some have been updated or slightly modified from the previous version having more details with the last developments and after the deployment and validation phase since most of them are related to the experimentations at pilots.

The OPTIMAI data management plan is a living document that has been regularly updated throughout the project as more information about the data within the project has become available and new datasets have emerged. This is the final version released at M42 (end of the project).

Table of Contents

LIST OF FIGURES

LIST OF TABLES

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.

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

Along these lines, the 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 has been collected and/or generated and the way in which it has been handled during and after the end of the project along with the standards applied to this end.
- Describe details on how the data has been 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 shows the datasets in use at M42. OPTIMAI takes a FAIR approach to data management following the European Commission's guidelines. Combined these two aspects comprise the overall OPTIMAI Data Management Plan.

1.1 Relation to other project work

This DMP has impact on the actions that has been performed over datasets that has been 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:

- 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 datasets that are 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

D1.5 is the fourth release of the OPTIMAI Data Management Plan (DMP), which updates and replaces the previous ones (D1.2, D1.3 and D1.4). 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, its results and with the EU guidelines about data management.

This section highlights the updates made in this fourth version D1.5 at M42 with regards to second version D1.4 delivered at M24:

- Updated Executive Summary to adapt the contents to the current version.
- Minor updates in Section 1.
- Minor changes in Section 2
- Minor changes in Section 3.
- Updated Section 4 with revised and new OPTIMAI datasets.
- Updated Conclusions.

2.Legal framework

The OPTIMAI approach has been 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 adhered 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: Everyone has the right to respect for his private and family life, his home and his correspondence.
- There shall be no interference by a public authority with the exercise of this right, excepting if it is in accordance with the law and democratic principles, as well as a necessary to protect 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 was processed it was:

- 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
- Ensured that data processing procedures for access to personal data were transparent [6].

Particular attention was 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) was undertaken during the course of the project, and consent was 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 upheld 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 were facilitated by the provision to data subjects of contact details of data controllers where research data collection is taking place, and by 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 are 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 upheld 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 adhered 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 follows 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 with 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, the FAIR data management principles focusing on how each general principle is specifically applicable in OPTIMAI and any specific processes and requirements to consider have been set out. Then a focus is given to personal data, which require special attention in data management. Finally, the key legal, ethical, data protection and security considerations for the management of data have been highlighted.

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 development.

The first stage of the data management process was to understand the broad context of the dataset in question; the datasets were summarised by answering the following six questions as set out in the European Commission's (2016) guidelines:

1. What is the purpose of the data collection/generation and its relation to the objectives of the project?

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

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')?

In this forth iteration of the DMP the previously identified datasets have been updated when needed, and further datasets that were not included in previous versions have been added.

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 (see Section 3.1.1 for details): it is 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. Point (5) concerning the size of the data is 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/Horizon Europe 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. In OPTIMAI we consider six main data types that can be used as umbrella terms within the project.

- **Project management data** – datasets 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** – 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)** – 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** –datasets created from the output of processing activity by a specific software module in OPTIMAI.
- **Publicly available dataset** – 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 –** any datasets 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.

No other data types emerged.

3.1.2 Data collection

This section presents the summary of datasets that have been collected and generated during the project, and that the partners have identified (detailed in Section 3.8).

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 have been 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.

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

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

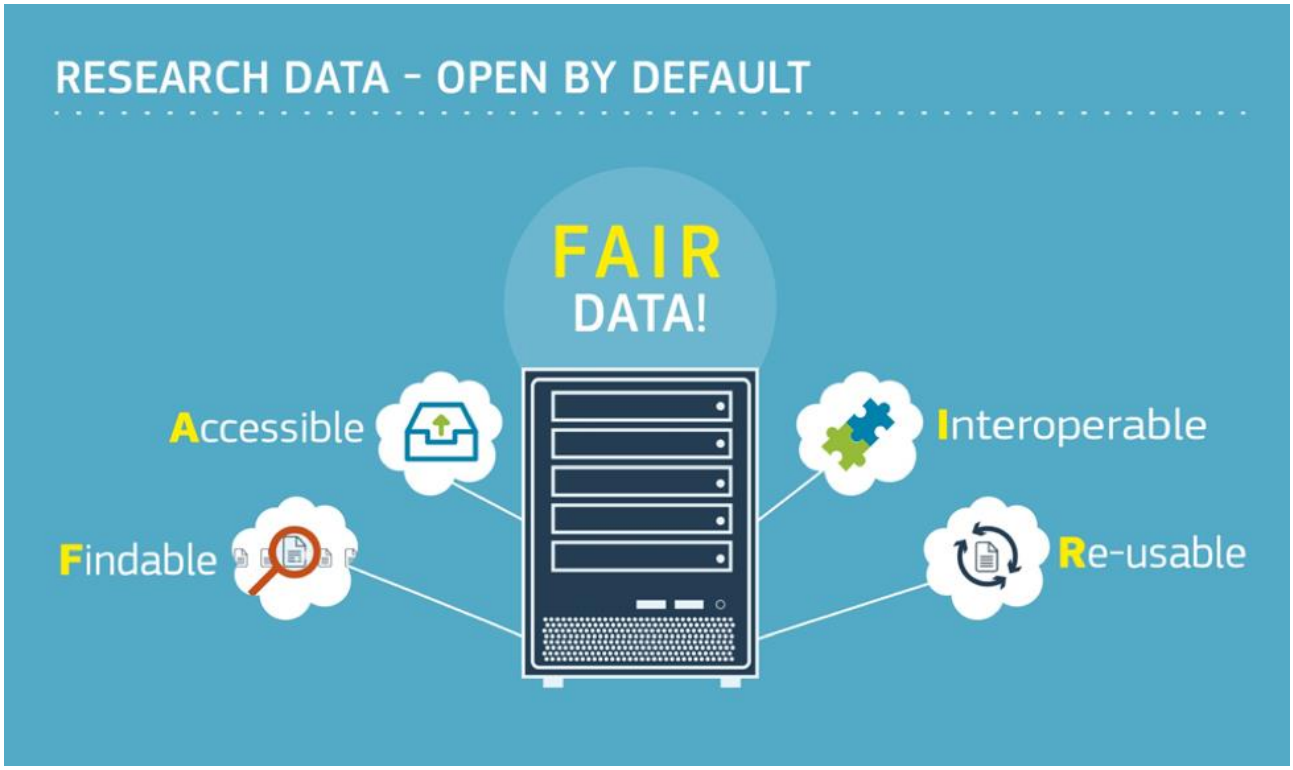

Figure 1: FAIR Data principles[5]

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 what FAIR data means in the context of OPTIMAI by considering these guidelines has been set out, the original description from Wilkinson et al., and the GO FAIR Initiative.

This was intended to be living documentation and the DMP has been 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 has been 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 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

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

'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.

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 has been 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 any 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.

Non-confidential research data that can be made openly available is deposited in either Zenodo, (OpenAIRE's trusted repository hosted by CERN), or in a trusted institutional repository, by the partner(s) working with such data.

Open-access scientific publications are deposited in a) Zenodo through CARR's H2020 account, or b) in a relevant institutional repository by the relevant author(s).

All publications uploaded to Zenodo are tagged so that the European Commission is visible as the funder and OPTIMAI and the Grant Agreement number 958264 are linked under "Award".

OpenAIRE (Zenodo) is directly linked to the European Commission's project database to enable cross-linking between OpenAIRE and CORDIS.

Public deliverables listed in the GA have been made publicly available via the project website.

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*).

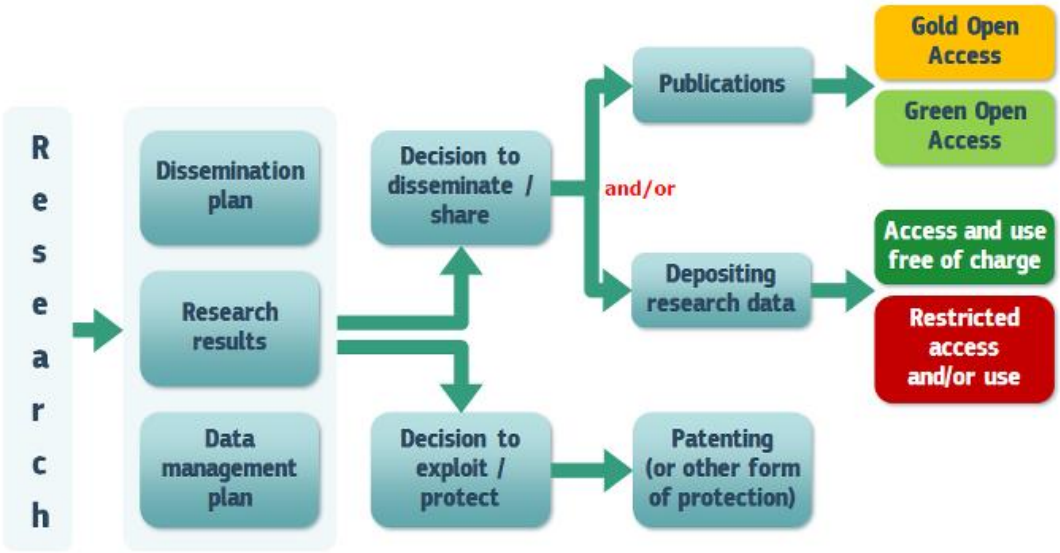

Figure 2: Approach for Open Access research data (H2020 Online Manual)[7]

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

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 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.

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 adoptes in its data management methodology the use of metadata vocabularies, standards, and methods that increases the interoperability of the data collected/generated through its activities.

More specifically, standard vocabularies have been used for all data types present in the project. For uncommon vocabulary, a clear mapping has been provided in order to facilitate its use. Thus, the project's data is interoperable and easy for sharing among researchers, institutions and organisations.

Partners observed OpenAIRE guidelines for online interoperability[8], furthermore they have also ensured that project data observed 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 produced a substantial volume of data and knowledge could 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 has an individual license. Access to project data is provided to the whole project Consortium and exclusively for the project objectives. Datasets produced as a result of the project work are shared within the Consortium and are only 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 could 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,
- a cookie ID,

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

- 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 followed these principles from initial collection through to long-term storage and archiving (or disposal) throughout the project. In particular, partners assured that they have established a lawful basis for any data they processed. 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 relied 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 collected and processed data only if, and insofar as, it was necessary for its research. Interviewees and validation test participants were asked for their signed informed consent and a clear description of the procedures that were used for data control and

anonymisation were 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

Over the course of the project, personal data was collected only insofar as it was 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 fell under two categories 1) research and 2) communication (dissemination and exploitation of results).

The legal bases for data collection for research activities was primarily 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 were 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 provided 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 were retained by data controllers in order to furnish as required by relevant authorities.[12]

If it was necessary to re-purpose existing datasets, it was 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 were deployed to protect personal data.[14]

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

Project partners did not currently envision processing any special categories of data[15]. However, in the event that this was necessary it was expected to 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]

---

[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)

For communication activities, the legal basis was also consent. Interested persons had the ability to subscribe to the OPTIMAI mailing list or newsletter by using the project website and were 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 was conducted ahead of sending any newsletter to persons to whom the partner's decide may be interested.[17] Such persons were free at any time to opt out of communications through clearly indicated means.

The rights of all data subjects were 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 was 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 was not necessary for the completion of tasks was destroyed as soon as task/activity will be completed.

Personal data is stored only for as long as necessary to achieve the goals and objectives of the project,[19] and stored personal data was 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 executed 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 were 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 made reasonable efforts to contact the concerned data subjects and provide them with the

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

required information. Appropriate safeguards were implemented where this was not possible, and the use of any such data were publicised on the project website.[20]

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 was 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 took 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 endeavoured 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 adhered to the requirements of Article 18(1)(a)-(d) and store the relevant data until the issue is resolved, whereby the personal data was 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 processed 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 obtained consent from data-subjects in advance of data collection and data-subjects were free to decline to provide consent or withdraw it at any time. Data-subjects were 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

[20] Article 14(5)(b)
[21] Article 17(3)(d)
[22] Article 22(2)(c), GDPR

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 exceptional restrictions were 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[24][25], 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 were informed about international data transfers at point of collection of personal data and were given the option of providing or declining their consent to the use of cookies where the project website is concerned. International services used were compliant with Article 46(2)(d), which is to say utilisation of these services were 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 collected and processed 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 was individually 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 established 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 could 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 were made transparent such as through the project website.[30]

Project partners were 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 complied 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 were 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 was 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 adhered 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 followed 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 acknowledged that responsibility applies to both the research and innovation (R&I) process as well as its outcomes. OPTIMAI's objective was 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 was devoted to the ethics and legal aspects of the project. The purpose of WP9 was to ensure that all research activities conducted in OPTIMAI were legally compliant, ethically acceptable and socially desirable. To that end, four tasks were foreseen: T9.1 focused on identifying the ethical and legal framework applicable to the project; T9.2 identified and analysed the ethical, legal and societal risks that the research activities may pose; T9.3 entailed the definition and implementation of the OPTIMAI Regulatory Model; lastly, T9.4 involved 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 reflected 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 submitted within WP9 are listed in the Table below (See Table 3). The involvement of the OPTIMAI Consortium as a whole was 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
M42	D9.4	Report on OPTIMAI ethical, legal and societal risks - 3rd version	TRI
M42	D9.8	Report on the OPTIMAI Regulatory Model - 4th version	UAB
M42	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 reflected 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.

As per the EB tasks described in the previous versions of the DMP, by M24 the EB has performed the following activities:

- The EB has reviewed all WP9 Deliverables submitted to date: D9.1 (Report on the OPTIMAI ethical and framework), D9.2 (Report on OPTIMAI ethical, legal and societal risks - 1st version), D9.3 (Report on OPTIMAI ethical, legal and societal risks – 2nd version), D9.4 (Report on OPTIMAI ethical, legal and societal risks – 3rd version), D9.5 (Report on the OPTIMAI Regulatory Model - 1st version), D9.6 (Report on the OPTIMAI Regulatory Model

– 2nd version) and D9.7 (Report on the OPTIMAI Regulatory Model – 3rd version) and D9.9 (Report on the ethical, legal and societal impacts of OPTIMAI) and has provided recommendations to ensure that the project's research activities are legally and ethically compliant and socially desirable.

- The EB has also reviewed D7.3 (Ethics recommendations and regulatory framework), in which recommendations were provided regarding the national legal frameworks of the countries in which the pilots will be conducted (UK, Spain and Greece). Recommendations covered data protection law, employment law, equality law, and health and safety law.
- Members of the EB have attended the Plenary Meetings, and the EB Chair has delivered presentations about the work conducted.

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 securely handled 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 occurs 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 ensures 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 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.

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 has been 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 measures have been undertaken by the partners responsibles of datasets to prevent data loss risks. In fact, confidentiality must be strictly maintained and anonymization has been applied, where required. In addition, partners did not allow to reveal sensitive data about participants to test cases, pilots, trials etc. These same principles have been taken into consideration in the dissemination of data.

3.7 Data archiving and preservation

The DMP provides 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 have to be updated each time they change in the project lifetime. The new updated dataset are identified by the number of the previous dataset version plus one, according to the naming convention reported above.

Project datasets for training tasks have been 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 have stored 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 the 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 appropriate resources and systems in place to share and preserve the data[35] are available.

So, a correct allocation of resources includes:

- estimating the costs for making the data FAIR and describing the methods of covering these costs;

[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

- 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: website hosting, €150/year (for 3 years + 5 years after the project) = €1,200, plus software (Adobe) €800 and plug-ins €400. The total amounts to €2,400.
- 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 is 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 that identified personal data contained in the datasets filled also 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 have been updated over time.

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 has been collected via suitable sensors in the production line of each pilot, the datasets have been mapped over the tasks.

Table 6: Overview of tasks and datasets in WP3

Task No	Task Name	Responsible	Related Datasets
T3.1	Multisensorial data acquisition and actuation network	EVT	WP3_4,
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	N/A
T3.4	On-the-edge processing for acquisition and actuation	ENG	N/A
T3.5	Blockchain framework for traceability and data integrity	CERTH	WP3_9 Blockchain
T3.6	AI components for quality control towards zero defect manufacturing	UTH	WP3_1 Defect_detection_TELEVES WP3_2 Defect_detection_KLEEMANN WP3_3 Defect_detection_MICROSEMI WP3_10 Image data for TVES UC1 WP3_11 Historical Time series data for KLEE sample 2 years WP3_12 Root_cause_samples

Overview	
Dataset ID	WP3_1
Dataset Title	Defect_detection_TELEVES
Work Package	WP3
Task / Deliverable	T3.6
Partner(s)	TELEVES
Data Type	jpg
Format	Not finalized yet
Details	
Description	A set of files pertaining to the products' inspection to detect defects. Collected in the production line via a machine vision RGB camera.
Data Size	Not finalized yet
Status	Established

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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	The data are available via the middleware API
Increase data re-use	*N/A*
Ethics and Data Protection	
Personal Data	Yes 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 collected only the necessary data for research aims, in the context of the project. CERTH applied 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 has been 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	T3.6
Partner(s)	KLEEMANN

Data Type	Derived
Format	json
Details	
Description	A set of files pertaining to the products' inspection to detect defects. Image data have been collected through the use of a wabcam. Elevator-weight displacement and velocity are estimated using a soft sensor model.
Data Size	Not finalized yet
Status	Established
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	Benchmarking for offline reinforcement learning algorithms.
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	With any programming language – machine learning framework
Increase data re-use	*N/A*
Ethics and Data Protection	
Personal Data	No personal data is included.
Security Requirements	CERTH collected only the necessary data for research aims, in the context of the project. CERTH applied 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 has been 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	T3.6
Partner(s)	MICROSEMI
Data Type	Derived
Format	json
Details	
Description	A set of files pertaining to the products' inspection to detect defects. Dataset files contain information about the volume of each glue deposit of a small number of sample PCBs. The dataset was created and annotated using a laser profilometer and a machine vision camera.
Data Size	1980 image samples, 4.9 MB
Status	Established
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	Benchmarking for robust regression algorithms under the influence of label noise
Open Data	
Is the data open?	Yes
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	With any programming language – machine/image processing framework

Increase data re-use	*N/A*
Ethics and Data Protection	
Personal Data	No personal data is included.
Security Requirements	CERTH collected only the necessary data for research aims, in the context of the project. CERTH applied 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 has been periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation..

Overview	
Dataset ID	WP3_4
Dataset Title	dimensional quality control
Work Package	WP3
Task / Deliverable	T3.1
Partner(s)	UNIMET, MTCL
Data Type	Derived
Format	.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	Established
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_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	Established
Use in OPTIMAI	The data have been 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 are 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_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	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

Overview	
Dataset ID	WP3_9

Dataset Title	Blockchain
Work Package	WP3
Task / Deliverable	T3.5
Partner(s)	All
Data Type	Data from Middleware
Format	json
Details	
Description	Data from Middleware, that are critical to be strored. Blockchain do not store any type of personal data.
Data Size	Not finilized yet
Status	Established
Use in OPTIMAI	The Blockchain is used to store critical data, for integrity, traceability and monitoring
Use beyond OPTIMAI	Because of the critical data, it is uncertain whether data can be used beyond OPTIMAI
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	OPTIMAI blockchain/ in a server inside CERTH facilities
Who	CERTH
Metadata	N/A
How	The data will be available via the middleware API
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	No personal data is included
Security Requirements	The blockchain can only be accessed from Middleware, with trusted credentials complying to the authorisation level of each authenticated requesting entity. Furthermore it is planned to have access control in the blockchain, and role based access control regarding the Smart Contracts.

Overview	
Dataset ID	WP3_10
Dataset Title	Image data for TVES UC1
Work Package	WP3, WP7
Task / Deliverable	T3.6
Partner(s)	UTH, TVES, CERTH, EVT
Data Type	Primary image Data (2D) collected by CERTH and EVT in OPTIMAI
Format	jpg, png, xml
Details	
Description	Images and labelling information that were collected from different Antenna samples that CERTH and EVT possess
Data Size	1.8 GB
Status	Data bases of images from the process.
Use in OPTIMAI	To locate and classify (through object detection algorithms) defects at antennas
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	-
Storage location	In Server inside UTH facilities
Who	UTH
Metadata	-
How	With any image processing program
Increase data re-use	
Ethics and Data Protection	
Personal Data	No
Security Requirements	No

Overview	
Dataset ID	WP3_11
Dataset Title	Historical Time series data for KLEE sample 2 years
Work Package	WP3, WP7
Task / Deliverable	T3.6, T5.4
Partner(s)	UTH, KLEE, FORTH
Data Type	Secondary historical Data collected by KLEE for a 2-year period
Format	csv
Details	
Description	Speed, pressure and noise time series data that were derived from the final power units quality tests in KLEE; plus a csv file that contains information about the parts that had been used in each clients order.
Data Size	26 MB
Status	already in place
Use in OPTIMAI	To estimate abnormal behaviours and perform anomaly detection in the processing variables. To train a recommendation service for advising the operators about the optimal velocity and pressure
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	-
Storage location	In Server inside UTH, FORTH facilities
Who	UTH, FORTH
Metadata	-
How	With Microsoft Excel and/or programming languages
Increase data re-use	-
Ethics and Data Protection	
Personal Data	No

Security Requirements	No

Overview	
Dataset ID	WP3_12
Dataset Title	Root_cause_samples
Work Package	WP3, WP7
Task / Deliverable	T3.6
Partner(s)	UTH, KLEE, CERTH
Data Type	Primary Time Series Data collected by CERTH and KLEE
Format	csv, json
Details	
Description	Speed, pressure and noise time series data that were derived from the auto-calibration tests which performed by CERTH in KLEE's Lab. It contains the actions that were performed by CERTH's auto-calibration system and the corresponding results in the monitoring Variables
Data Size	20 MB
Status	already in place
Use in OPTIMAI	To estimate abnormal behaviours and perform Root Cause Analysis
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	-
Storage location	In Server inside UTH facilities
Who	UTH
Metadata	-
How	With Microsoft Excel and/or programming languages
Increase data re-use	-

Ethics and Data Protection	
Personal Data	No
Security Requirements	No

4.4 WP4 - Production virtualization

WP4 targets the development of software components for production planning and optimization with zero defects and downtime. Different solutions around virtualization, digital twinning, and AI have been developed until the end of the project (M42) in WP4's tasks. The datasets and their utilization are detailed 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	WP4_1 eCat, WP4_4 Production datasets, WP4_5 OPTIMAI Simulation Library
T4.2	Sensors modelling and virtualization	FINT	N/A
T4.3	Prediction of upcoming defects towards zero downtime manufacturing	CERTH	WP4_3 KLEE: Time-series data from elevator movement tests TVES: Images of defective/non-defective antennas. Time-series data that were collected from FINOT middleware were not used. MTCL: Images from PCBs collected from the middleware and internal partners WP4_4 TVES: FEM simulation videos of antenna bending. WP4_5 MTCL Deionizaton Water Plant, WP4_6 Production datasets,

			WP4_7 OPTIMAI Simulation Library
T4.4	Simulation engine for production planning	VIS	WP4_1 eCat

Overview	
Dataset ID	WP4_1
Dataset Title	eCat
Work Package	WP4
Task / Deliverable	T4.1/D4.4, T4.4/D4.8
Partner(s)	VIS
Data Type	Virtual simulation components to be used within the simulation/virtualization and to extend into the digital twin.
Format	.vcmx, .vcm
Details	
Description	Virtual simulation components libary developed, and maintained by VIS and contain the following datasets: Geometries (CAD datasets), Behaviours, and Parameters (design and operation datasets).
Data Size	~2GB
Status	In place, collected by VIS
Use in OPTIMAI	The available simulation components will be the starting point for creating simulation models and virtualizing the production scenarios.
Use beyond OPTIMAI	The simulation components available in the eCat are currently used within different research and commercial projects. It is expected that simulation components developed within OPTIMAI, will become part of the eCat after the end of the project, after going through the internal release process.
Open Data	
Is the data open?	No
Explanation	
Storage location	

Who	Visual Components Oy
Metadata	Production metadata is available
How	Metadata only accessible through Visual Components 4.0 UI.
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	No
Security Requirements	
Comments	

Overview	
Dataset ID	WP4_3
Dataset Title	KLEE: Time-series data from elevator movement tests TVES: Images of defective/non-defective antennas. Time-series data that were collected from FINT middleware. MTCL: Images from PCBs collected from the middleware and internal partners
Work Package	WP4
Task / Deliverable	T4.3, D4.2, D4.5, D4.6
Partner(s)	CERTH
Data Type	collected by CERTH, TVES, KLEE
Format	csv, png, xlsx
Details	
Description	
Data Size	KLEE: Time series data containing 7 columns of elevator movement measurements such as NOISE_UP, SPEED_UP, PRESSURE_UP and the respective DOWN movement measurements. The dataset consists of 109,971 rows and 4.2MB. 1) TVES: 1) The initial dataset consisted of 20 real production line samples with defects: 19 showing Folding Error (FERR) defects, and 1 displaying Broken Rod (BROD) and Broken End (BEND) defects. Using Generative Adversarial Networks, the dataset was expanded to include 355 new synthetic samples, adding over 100

	instances of FERR and more than 150 instances of BROD and BEND defects. Additionally, 108 working samples were gathered from the production line. 2) Time-series data taken from FINOT containing no. of defective antennas produced, error and working times, and so on. The original dataset consisted of 1,734,330 rows (from 09/2022 to 09/2023) and 323MB. This dataset was updated with more recent data, extending to 2,839,858 rows (from 09/2022 to 03/2024) and 558MB. MTCL: 1) A single PCB photographed from different angles. 2) A single PCB rotated in a 3D processing software (Blender)
Status	KLEE: in place TVES: Synthesis in place, defect forecasting/prediction in place MTCL: Synthesis in place
Use in OPTIMAI	Synthetic image and data generation / future defect prediction
Use beyond OPTIMAI	N/A
Open Data	No
Is the data open?	No
Explanation	
Storage location	In a server inside CERTH facilities
Who	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Requesting for it manually or through FINOT middleware
Increase data re-use	-
Ethics and Data Protection	
Personal Data	No
Security Requirements	No

Overview	
Dataset ID	WP4_4
Dataset Title	TVES: FEM simulation videos of antenna bending.
Work Package	WP4

Task / Deliverable	T4.3, D4.6
Partner(s)	CERTH
Data Type	Sythetic/generated data
Format	mp4
Details	
Description	28 scenarios/videos of antenna bending based on various misplacements on the bending surface, utilizing the Finite Element Method. Data were created using the ANSYS software.
Data Size	11.6 MB
Status	Specify if it is already in place, established, or planned??
Use in OPTIMAI	Defect prediction
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	No
Security Requirements	No

Overview	
Dataset ID	WP4_5
Dataset Title	MTCL: Deionizaton Water Plant
Work Package	WP4

Task / Deliverable	T4.3, D4.6
Partner(s)	CERTH
Data Type	Primary data collected by MTCL and synthetic data by CERTH
Format	csv
Details	
Description	The dataset was collected from the deionization water plant in the wafer sawing process. This dataset was expanded using simulated data.
Data Size	The dataset consists of 3 variables: i) resistivity of the primary filter, ii) resistivity of the secondary filter and iii) power_reading. The original dataset spanned 2 weeks (from 29/01/2024 to 13/02/2024) and 4,9MB. The dataset was first extended using simulated data of 5 months and then further extended using another 5 months of real-production data.
Status	In place
Use in OPTIMAI	The data will be used for defect forecasting/prediction
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No
Explanation	N/A
Storage location	In a server inside CERTH facilities
Who	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Requesting for it manually or through FINOT middleware
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	No
Security Requirements	No

Overview	
Dataset ID	WP4_6
Dataset Title	Production data TVES: Antenna manufacturing MTCL: PCB manufacturing
Work Package	WP4
Task / Deliverable	T4.1/D4.2, T4.4/D4.8
Partner(s)	VIS, TVES, MTCL
Data Type	Primary data collected by the partners for equipment, products, and processes, used to model and configure the simulation components,
Format	.txt, .csv, .xlsx, .docx, .doc, .gcode, .vcmx, .vcm, .json, .stp and CAD (several formats)
Details	
Description	Production datasets refer to information required to configure a manufacturing system (machine or set of equipment) to manufacture a production order. These datasets can include production orders, product recipes, CAD, Product Management Information (PMI), equipment configuration parameters, robot programs, and Gcode.
Data Size	~2GB
Status	In place
Use in OPTIMAI	The production datasets have been used to virtualize manufacturing systems
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No.
Explanation	
Storage location	Visualcomponents.sharefile.com/ (private project repository per pilot for use within the project activities)
Who	Visual Components Oy (for use within the project activities)

Metadata	Yes, Metadata is generated for manufacturing processes.
How	Metadata generated from the production datasets through the simulation accessible through Visual Components 4.0 GUI
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	No personal data is included
Security Requirements	The datasets are located on a secure server with double-factor authentication. VIS maintains their storage during the project.

Overview	
Dataset ID	WP4_7
Dataset Title	OPTIMAI Simulation Library
Work Package	WP4
Task / Deliverable	T4.1/D4.2, T4.4/D4.3
Partner(s)	VIS
Data Type	Primary data collected by VIS
Format	vcmx, .vcm,
Details	
Description	OPTIMAI simulation library contains the components developed by VIS for OPTIMAI
Data Size	~2GB
Status	In place
Use in OPTIMAI	The developed simulation components are used to virtualize the OPTIMAI pilot lines.
Use beyond OPTIMAI	N/A
Open Data	
Is the data open?	No.
Explanation	

Storage location	Visualcomponents.sharefile.com/ (private project repository per pilot for use within the project activities)
Who	Visual Components Oy (for use within the project activities)
Metadata	Yes, metadata generated with production processes
How	Metadata generated from the simulation components using the simulation is accessible through Visual Components 4.0 GUI
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	No personal data is included
Security Requirements	The datasets are located on a secure server with double-factor authentication. VIS maintains their storage during the project.

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 are collected. These datasets do not 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 have been adopted.

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 WP5_11 Gesture_recognition_KLEEMANN

			WP5_12 Gesture_recognition_TELEVES WP5_13 Gesture_recognition_MICROSEMI
T5.1	Perception techniques for operator-machine interaction	CERTH	WP5_5 instance segmentation_KLEEMANN WP5_6 instance segmentation_TELEVES WP5_7 instance segmentation_MICROSEMI WP5_8 Pose_estimation_KLEEMANN WP5_9 Pose_estimation_TELEVES WP5_10 Pose_estimation_MICROSEMI WP5_14 EgoGesture Dataset WP5_15 SyntheticEgoGesture Dataset
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	Not yet finalized
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. An activity recognition model will be trained in order to understand the operator's task at a given time and provide the contextual information to DSS based on the operator's viewpoint and current activity.
Data Size	Not yet finalized
Status	Established
Use in OPTIMAI	To support AR functionalities (UC 2 - Production line setup-calibration). Specifically, for TELEVES speed up operators 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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	Yes, 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	Not yet finalized
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. An activity recognition model will be trained in order to understand the operator's task at a given time and provide the contextual information to DSS based on the operator's viewpoint and current activity.
Data Size	Not yet finalized
Status	Established
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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	Yes, 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	
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. An activity recognition model will be trained in order to understand the operator's task at a given time and provide the contextual information to DSS based on the operator's viewpoint and current activity.
Data Size	
Status	Established
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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models
Increase data re-use	N/A
Ethics and Data Protection	
Personal Data	Yes 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	Partners data for project communication. Data were collected with direct email communication.
Data Size	NA
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	CEO - Pietro Carratù
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.

Overview	
Dataset ID	WP5_5
Dataset Title	instance segmentation_KLEEMANN
Work Package	WP5
Task / Deliverable	T5.1
Partner(s)	KLEEMANN
Data Type	Derived, Synthetic
Format	png, avi
Details	
Description	A set of real and synthetic images will be used to train an instance segmentation model. The set of images contains the valve block and its individual parts of interest (valves, oil pumb etc.). 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_KLEEMAN dataset, metrological measurements, operator's heart rate etc.) will be executed and propagated to the foreseen AR environment.
Data Size	Real images=594, Synthetic images=7200, Total=7794
Status	Established
Use in OPTIMAI	To train an NN model 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	Nikolaos Dimitriou
Metadata	N/A

How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	Yes, the dataset includes personal data of the operators. Specifically, includes images of the valve block, but not the operators. There is no need or intention to process and store identification data like name, surname, contact details etc.
Security Requirements	CERTH collected only the necessary data for research aims, in the context of the project. CERTH applied 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 has been periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP5_6
Dataset Title	instance segmentation_TELEVES
Work Package	WP5
Task / Deliverable	T5.1
Partner(s)	TELEVES
Data Type	Derived, Synthetic
Format	png, avi
Details	
Description	A set of real and synthetic images will be used to train an instance segmentation model. The set of images contains antennas and its individual parts of interest. 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. 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	Real images=313, Synthetic images=7000, Total=7313
Status	Established
Use in OPTIMAI	To train an NN model to support AR functionalities (UC 2 - Production line setup-calibration). Specifically, for TELEVES speed up operators 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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	Yes, the dataset includes personal data of the operators. Specifically, includes images of the antennas, but not the operators. There is no need or intention to process and store identification data like name, surname, contact details etc.
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_7
Dataset Title	instance segmentation_MICROSEMI
Work Package	WP5

Task / Deliverable	T5.1
Partner(s)	MICROSEMI
Data Type	Derived, Synthetic
Format	png, avi
Details	
Description	A set of real and synthetic images will be used to train an instance segmentation model. The set of images contains 2 subsets of different PCBs. 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. 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	Real images= 1600, Synthetic images=6000, Total=7600
Status	Established
Use in OPTIMAI	To train an NN model 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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	Yes, the dataset includes personal data of the operators. Specifically, includes images of the wafers and PCBs, but not the

	operators. There is no need or intention to process and store identification data like name, surname, contact details etc.
Security Requirements	CERTH collected only the necessary data for research aims, in the context of the project. CERTH applied 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 has been periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP5_8
Dataset Title	Pose_estimation_KLEEMANN
Work Package	WP5
Task / Deliverable	T5.1
Partner(s)	KLEEMANN
Data Type	Synthetic
Format	png, avi
Details	
Description	png, avi
Data Size	png, avi
Status	Established
Use in OPTIMAI	To train an NN model 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	Nikolaos Dimitriou (CERTH)

Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	Yes, the dataset includes personal data of the operators. Specifically, includes images of the valve block, but not the operators. There is no need or intention to process and store identification data like name, surname, contact details etc.
Security Requirements	CERTH collected only the necessary data for research aims, in the context of the project. CERTH applied 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 has been periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP5_9
Dataset Title	Pose_estimation_TELEVES
Work Package	WP5
Task / Deliverable	T5.1
Partner(s)	TELEVES
Data Type	Synthetic
Format	png, avi
Details	
Description	A set of real and, mostly, synthetic images will be used to train a pose estimation model. The set of images contain antennas. After the instance segmentation algorithm applied, the objects' relative position with respect to a human operator will be recognised via the AR glasses.
Data Size	Real images=474
Status	Established

Use in OPTIMAI	To train an NN model to support AR functionalities (UC 2 - Production line setup-calibration). Specifically, for TELEVES speed up operators 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	Nikolaos Dimitriou
Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	Yes, the dataset includes personal data of the operators. Specifically, includes images of the antennas, but not the operators. There is no need or intention to process and store identification data like name, surname, contact details etc.
Security Requirements	CERTH collected only the necessary data for research aims, in the context of the project. CERTH applied 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 has been periodically revised in order to incorporate technological novelties and updating methods in alignment with GDPR regulation.

Overview	
Dataset ID	WP5_10
Dataset Title	Pose_estimation_MICROSEMI
Work Package	WP5
Task / Deliverable	T5.1
Partner(s)	MICROSEMI

Data Type	Synthetic
Format	png, avi
Details	
Description	A set of real and, mostly, synthetic images will be used to train a pose estimation model. The set of images contain wafers and PCBs. After the instance segmentation algorithm applied, the objects' relative position with respect to a human operator will be recognised via the AR glasses.
Data Size	png, avi
Status	Established
Use in OPTIMAI	To train an NN model 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	Nikolaos Dimitriou
Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	Yes, the dataset includes personal data of the operators. Specifically, includes images of the wafers and PCBs, but not the operators. There is no need or intention to process and store identification data like name, surname, contact details etc.
Security Requirements	CERTH collected only the necessary data for research aims, in the context of the project. CERTH applied 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 has been periodically revised

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

Overview	
Dataset ID	WP5_11
Dataset Title	Gesture_recognition_KLEEMANN
Work Package	WP5
Task / Deliverable	T5.1
Partner(s)	KLEEMAN
Data Type	Synthetic, Publicly available
Format	Not finalized yet
Details	
Description	A set of images (synthetic, opensource) will be used to train a gesture recognition model. The set of images contain gestures that are described in the gesture vocabulary, defined in the framework of OPTIMAI. The model will recognize the operator's gestures via the AR field of view in order to adjust parameters in the machinery.
Data Size	Not finilized yet
Status	Established
Use in OPTIMAI	To train an NN model 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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models

Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	No, specific personal data are requested
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_12
Dataset Title	Gesture_recognition_TELEVES
Work Package	WP5
Task / Deliverable	
Partner(s)	TELEVES
Data Type	Synthetic, Publicly available
Format	Not finalized yet
Details	
Description	A set of images (synthetic, opensource) will be used to train a gesture recognition model. The set of images contain gestures that are described in the gesture vocabulary, defined in the framework of OPTIMAI. The model will recognize the operator's gestures via the AR field of view in order to adjust parameters in the machinery.
Data Size	Not finilized yet
Status	Established
Use in OPTIMAI	To train an NN model to support AR functionalities (UC 2 - Production line setup-calibration). Specifically for TELEVES speed up operators 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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	No, specific personal data are requested
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_13
Dataset Title	Gesture_recognition_ MICROSEMI
Work Package	WP5
Task / Deliverable	
Partner(s)	MICROSEMI
Data Type	Synthetic, Publicly available
Format	Not finalized yet
Details	
Description	A set of images (synthetic, opensource) will be used to train a gesture recognition model. The set of images contain gestures that are described in the gesture vocabulary, defined in the framework of OPTIMAI. The model will recognize the operator's

	gestures via the AR field of view in order to adjust parameters in the machinery.
Data Size	Not finilized yet
Status	Established
Use in OPTIMAI	To train an NN model 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	Nikolaos Dimitriou (CERTH)
Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	No, specific personal data are requested
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_14
Dataset Title	EgoGesture Dataset
Work Package	WP5
Task / Deliverable	T5.1

Partner(s)	YBQ
Data Type	Publicly available dataset
Format	avi, jpg, csv
Details	
Description	A multi-modal (RGB and Depth) large scale dataset for egocentric hand gesture recognition. The dataset contains 2,081 RGB-D videos, 24,161 gesture samples and 2,953,224 frames from 50 distinct subjects in 6 different indoor and outdoor scenes. It consists of 83 classes of static or dynamic gestures focused on interacting with wearable devices.
Data Size	Videos 46 GB, Images 32 GB
Status	In place
Use in OPTIMAI	Hand Gestures Recognition
Use beyond OPTIMAI	Hand Gestures Recognition
Open Data	
Is the data open?	Yes - but restricted access
Explanation	To obtain the database, an NDA need to be signed and sent to the owner. The database is released for research and educational purposes. All rights of the EgoGesture Database are reserved.
Storage location	After the download, the database is stored in the YBQ server.
Who	Pietro Carratù (CEO)
Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	No
Security Requirements	No

Overview	
Dataset ID	WP5_15

Dataset Title	SyntheticEgoGesture Dataset
Work Package	WP5
Task / Deliverable	T5.1
Partner(s)	YBQ
Data Type	Synthetic/generated - Primary data collected by YBQ in OPTIMAI
Format	avi, jpg, csv
Details	
Description	A multi-modal (RGB and Depth) large scale dataset for egocentric hand gesture recognition. The dataset contains RGB-D videos of static or dynamic gestures reproduced synthetically starting from the video acquisition of real gestures. The number of gesture classes and the total number of samples have not yet been established. It is focused on interacting with wearable devices.
Data Size	450 GB of images
Status	In place
Use in OPTIMAI	Hand Gestures Recognition
Use beyond OPTIMAI	Hand Gestures Recognition
Open Data	
Is the data open?	No
Explanation	
Storage location	Youbiquo servers
Who	CEO - Pietro Carratù
Metadata	N/A
How	Software, CNN models
Increase data re-use	TBD
Ethics and Data Protection	
Personal Data	No
Security Requirements	No

4.6 WP6 - Decision support and system integration

Within WP6 the OPTIMAI Decision Support System for early notifications regarding defects has been developed. Furthermore, WP6 provides the integrated solution and prepares the ground for realistic system testing, user engagement and evaluation of use-cases. The previously foreseen FINT dataset "WP6_1 Marketplace data" has been replaced by WP6_3 and WP6_4 after an agreement between the two parties on the development of the marketplace.

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	WP6_2 Personal credentials, Image and json data for DSS Webpages
T6.2	Intelligent Marketplace for AI sharing and scrap re-use	FINT	~~WP6_1 Marketplace data~~ WP6_3 Marketplace accounts WP6_4 Marketplace data
T6.3	Testing and incremental integration of components	ENG	N/A
T6.4	System validation	FORTH	N/A

Overview	
Dataset ID	WP6_2
Dataset Title	Personal credentials, Image and json data for DSS Webpages
Work Package	WP6
Task / Deliverable	T6.1
Partner(s)	CERTH, FINT
Data Type	Primary data collected by partner in OPTIMAI. Publicly available dataset (e.g. training / benchmark data)
Format	json
Details	

Description	The .jpg and .json files are collected by the middleware, stored in a database, and after processing produce updated versions that are stored back in the database. The entire historical timeline and statistics results are displayed on the DSS Webpage.For the credentials, the system will get them and store them in the database as soon as the user creates them
Data Size	1MB
Status	Already in place
Use in OPTIMAI	The DSS user is able to edit the data in a visual environment and produce updated versions of it. He will also be able to see the processing history as well as statistical results. Credentials will be used for each user to have access to their own files and data.
Use beyond OPTIMAI	Users are able to visually see the initial results. Then they will be able to make decisions in timely manner to prevent suboptimal operations and defect generations /propagations, and more time consuming decision for better production planning towards near zero manufacturing. Credentials are of no use to third party users
Open Data	
Is the data open?	No
Explanation	
Storage location	in a server inside CERTH facilities with limited access to authorized personal
Who	CERTH
Metadata	
How	The data is available via the Middleware
Increase data re-use	-
Ethics and Data Protection	
Personal Data	There is no personal information in the files. Only the credentials of each user are their username and password. If personal data needs to be processed, the system must be updated.
Security Requirements	Personal data as well as files will be secure with appropriate tools and will not be accessible to any third party

Overview	
Dataset ID	WP6_3
Dataset Title	Marketplace accounts
Work Package	WP6
Task / Deliverable	T6.2
Partner(s)	ENG
Data Type	Primary data
Format	json, csv
Details	
Description	Sign up form to get credentials to access the marketplace
Data Size	around 50MB
Status	in place
Use in OPTIMAI	There is not a direct use of these data, they just serve to allow the access to the marketplace
Use beyond OPTIMAI	These data will not be used after the project
Open Data	
Is the data open?	No
Explanation	
Storage location	ENG servers
Who	ENG
Metadata	No metadata associated
How	Only ENG has visibility on these data
Increase data re-use	No reuse is foreseen.
Ethics and Data Protection	
Personal Data	Yes, the dataset will include some details of the user (name, surname, organization, email address). The marketplace will have a banner with a link to the privacy and policy applied.

Security Requirements	The marketplace can only be accessed with trusted credentials complying to the authorisation level of each authenticated requesting entity. Furtermore the interactions with the marketplace are allowed only via TLS based encrypted communication channels and OIDC protocol.

Overview	
Dataset ID	WP6_4
Dataset Title	Marketplace data
Work Package	WP6
Task / Deliverable	T6.2
Partner(s)	ENG
Data Type	Derived data/primary data
Format	json, csv, jpeg, png
Details	
Description	Data describing a) the offered AI algorithms developed within OPTIMAI and b) the defective parts of an Industry
Data Size	from MB to 10GB
Status	in place
Use in OPTIMAI	AI algorithms offered to third parties, enhance scrap reuse reducing waste
Use beyond OPTIMAI	AI algorithm may serve as enabler to develop further Ucs by other parties.
Open Data	
Is the data open?	No
Explanation	-
Storage location	OPTIMAI marketplace
Who	ENG for the primary collected, other OPTIMAI components for those derived
Metadata	A series of mandatory tags has/have to be added to make the data findable.

How	The data will be available via the marketplace API and/or dashboard
Increase data re-use	Data are available at the mkpl url, visibility depends on user profile
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 these context, the suggested techniques recommended in the context of WP9, has been applied.
Security Requirements	The marketplace can only be accessed with trusted credentials complying to the authorisation level of each authenticated requesting entity. Furtermore the interactions with the marketplace allowed only via TLS based encrypted communication channels and OIDC protocol.

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 have been identified.

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 WP7_2 Image data for Kleemann use case 1
T7.4	Demo in TELEVES	TVES	WP7_3 Image/3D data for Telves use case 1:

			Reduce the number of defects
T7.5	Demo in MICROSEMI	MTCL	WP7_4 Sensor Datasets WP7_5 Image data for MTCL use case

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	Established
Use in OPTIMAI	The data are used mainly for quality inspection and early defect detection. The dataset assists the (re)configuration of the hydraulic units parameters and the production planning.
Use beyond OPTIMAI	The dataset is useful for other researchers and also end-users, since it will combine digital twins with AI models. Based on the above combination, the data provides 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

Overview	
Dataset ID	WP7_2
Dataset Title	Image data for Kleemann use case 1
Work Package	WP3, WP7
Task / Deliverable	T3.2, T7.3
Partner(s)	EVT, KLEEMANN
Data Type	Image data used for prototyping a sensor
Format	images, text
Details	
Description	Images and labeling information for object detection and segmentation related to the valve inspection at KLEEMANN..
Data Size	2GB
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 usecases

Storage location	The data has been manually provided by project partner CERTH and KLEEMANN. EVT respects the project partner's conventions and requirements. EVT is storing the data internally for the lifetime of the project.
Who	EVT and the project partners have shared the data.
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	WP7_3
Dataset Title	2D/3D data for Telves use case 1: Reduce the number of defects
Work Package	WP3, WP7
Task / Deliverable	T3.2, T7.4
Partner(s)	EVT, TELVES
Data Type	Image data used for prototyping a sensor
Format	images, text
Details	
Description	Images of the line scan camera
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 reated 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 usecases
Storage location	The data has been manually provided by project partners. EVT respects the project partner's conventions and requirements.
Who	Televes was sending parts to EVT, EVT generated the data within a preliminary evaluation. After installation of the sensor on production site, the data has been generated directly in the product line. Televes is sharing the data with EVT.
Metadata	EVT has used the data only for internal use, such as training deep learning models.
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	WP7_4
Dataset Title	Image/3D data for MTCL use case 1 (Defect detection) and Use case 2:(Process calibration)
Work Package	WP3, WP7
Task / Deliverable	T3.2, T7.5
Partner(s)	EVT, CERTH, FORTH, UAB, UNIMET, ENG, FINT, VIS
Data Type	Image data used for prototyping a sensor

Format	Images in various formats including .brv. jpg .png
Details	
Description	Images of Electronic products to determine previous manufacturing process output
Data Size	1 .brv file is typically 23Mb and the dataset for the dispense process (100 images) 2.3Gb Point clouds are even larger
Status	Data bases containing image data from the processes have been in continuous use.
Use in OPTIMAI	Used for manual /semi-automatic or automatic configuration and training of the process and analysis of the approaches and their suitability for the end user.
Use beyond OPTIMAI	The data is related to the specific pilot use cases and belongs to the pilot. MTCL may share the datasets on request but only with the guarantee of anonymity for the source.
Open Data	
Is the data open?	No
Explanation	Process Data comes from the OPTIMAI solutions but due to MTCL IT restrictions data has been shared with technical providers by sharing via Microsoft OneDrive. Other datasets have been generated locally at the technical providers where MTCL has provided samples of processed parts.
Storage location	Throughout the project, there have been lots of datasets related to the MTCL use cases, MTCL will retain copies of the process data for technical partners that require it at MTCL. Other datasets generated by technical partners have been stored locally at the technical provider.
Who	MTCL has shared data with EVT, CERTH, FORTH, UAB, UNIMET, ENG, FINT, VIS.
Metadata	MTCL does not expect the use of Metadata within the project.
How	Image data can be accessed with any image processing program, 3D data can be accessed with tools of EVT and UNIMET but may also be converted into other formats.
Increase data re-use	Not yet planned but Data has been generated in such a way it may be possible to re-use the data after the end of the Project.

Ethics and Data Protection	
Personal Data	No
Security Requirements	No

Overview	
Dataset ID	WP7_5
Dataset Title	Sensor data for MTCL use case 1 and 2: Defect detection and Process calibration
Work Package	WP3, WP7
Task / Deliverable	T3.2, T7.5
Partner(s)	EVT, CERTH, FORTH, UAB, UNIMET, ENG, FINT, VIS
Data Type	Sensor data used for generation of OPTIMAI solutions
Format	numeric, text typically .csv files
Details	
Description	Process Sensor data to be used in conjunction with image data with the aim of giving technical providers sufficient information to generate OPTIMAI solutions.
Data Size	During the course of the project the file sizes have varied from kb to 20Mb
Status	Data bases containing sensor data from the processes have been in continuous use.
Use in OPTIMAI	Used for manual /semi-automatic or automatic configuration and training of the process and analysis of the approaches and their suitability for the end user.
Use beyond OPTIMAI	The data is related to the specific pilot use cases and belongs to the pilot. MTCL may share the datasets on request but only with the guarantee of anonymity for the source.
Open Data	
Is the data open?	No
Explanation	Process Data comes from the OPTIMAI solutions but due to MTCL IT restrictions data has been shared with technical providers by

	sharing via Microsoft OneDrive. Other datasets have been generated locally at the technical providers where MTCL has provided samples of processed parts.
Storage location	Throughout the project, there have been lots of datasets related to the MTCL use cases, MTCL will retain copies of the process data for technical partners that require it at MTCL. Other datasets generated by technical partners have been stored locally at the technical provider.
Who	MTCL has shared data with EVT, CERTH, FORTH, UAB, UNIMET, ENG, FINT, VIS.
Metadata	MTCL does not expect the use of Metadata within the project.
How	Sensor data can be accessed with any spreadsheet type software package, but may also be converted into other formats for example .JMP used for statistical analysis.
Increase data re-use	Not yet planned but Data has been generated in such a way it may be possible to re-use the data after the end of the Project.
Ethics and Data Protection	
Personal Data	No
Security Requirements	No

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.

In previous versions, a dataset containing the newsletter recipients was foreseen. In the meanwhile a different choice was made: instead of using the Mailchimp platform as planned, the choice turned on LinkedIn newsletters. As a result, data on OPTIMAI newsletter recipients are not collected. Instead, subscribers receive a notification and an email when a newsletter is published. Hence, any personal data is collected (i.e., there is no visibility on email addresses etc. connected to the LinkedIn profiles of subscribers).

One internal dataset collects the exploitation contacts will be used for managing and communication purposes. Moreover, a dataset for key exploitable results has been established and updated till M42.

Table 11: Overview of tasks and datasets in WP8

Task No	Task Name	Responsible	Related Datasets
T8.1	High impact dissemination & communication activities	CARR	
T8.2	Clustering, networking & Knowledge transfer activities	CARR	WP8_2 Cluster mailing list
T8.3	Standardization activities	EVT	N/A
T8.4	Market analysis and segmentation	FINT	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	Cluster mailing list containing names and email addresses of cluster contacts who have opted in
Dataset ID	WP8_2
Dataset Title	Cluster mailing list
Work Package	WP8
Task / Deliverable	T8.2
Partner(s)	CARR
Data Type	Personal details collected by CARR during the OPTIMAI project activities
Format	xls/docx
Details	
Description	A list of names and email addresses of contact persons from the related projects. The details are stored in a Google sheet and the cluster partners have access to the document link.
Data Size	25 KB
Status	Established
Use in OPTIMAI	To liaise with relevant partners from other projects to network, create links and synergies and facilitate communication between projects.

Use beyond OPTIMAI	*N/A*
Open Data	
Is the data open?	No, only accessible to partners with the exact document link
Explanation	The dataset has been used to communicate with the 4ZDM cluster members.
Storage location	Google doc
Who	CARR updated and maintained the mailing list as needed.
Metadata	N/A
How	Through a direct link.
Increase data re-use	N/A.
Ethics and Data Protection	
Personal Data	Names and email addresses.
Security Requirements	Access is provided through a direct link (not open to the public).

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	Established

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 Nextcloud 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

D1.5 Data Management Plan 4th version, reports and updates 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 adheres to the FAIR principles and the ways that data have been 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 collected/generated under the project activities, methodologies to make dataset FAIR, the allocation of resources and the data security. Moreover, the procedures for storage and preservation of dataset and the ethical aspects addressed during the project are described. 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 has been an ongoing process that have been reviewed and updated throughout the project's timeline,.

6.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

7.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:.....

8.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].

9.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*

10. 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?*