iQonic

IQONIC

Innovative strategies, sensing and process chains for increased Quality, re-configurability, and recyclability of Manufacturing Optolectronics

Deliverable

D9.4 Data Management Plan

Deliverable Lead: ATLANTIS

Deliverable due date: 31/03/2019 (M06)

Actual submission date: 15/04/2019

Version: 1.0

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 820677

Document Control Page	
Title	D9.4 Data Management Plan
Creator	ATLANTIS
Description	This deliverable will describe how acquired data and knowledge will be shared, made open, maintained and preserved as well.
Contributors	FRAUHOFER, BRUNEL, CORE, POLIMI, SHADOW, HOLONIX, SENSAP, IP-ASCR (HiLASE), PRIMA, ALPES, FORTH, FILAR, SACMI, FICONTEC, TAU, BRIGHTERWAVE.
Creation date	20/11/2018
Type	Report
Language	English
Audience	☒ public ☐ confidential
Review status	☐ Draft ☒ WP leader accepted ☒ Coordinator accepted
Action requested	☐ to be revised by Partners ☐ for approval by the WP leader ☐ for approval by the Project Coordinator ☒ for acknowledgement by Partners

Table of Contents

1 Objectives 6
2 Introduction 7
2.1 Purpose of the data collection/generation and its relation to the objectives of the project 7
2.2 Types and formats of data the project will generate/collect 9
2.3 Existing data, re-use and how? 10
2.4 Expected size of the data 11
2.5 Data utility 11
3 FAIR data 12
3.1 Making data findable, including provisions for metadata 12
3.1.1 Metadata provision 12
3.1.2 Standards for metadata creation 13
3.1.3 Approach towards search keyword 13
3.1.4 Use of persistent and unique identifiers 14
3.1.5 Naming conventions 14
3.1.6 Clear versioning 14
3.2 Making data openly accessible: 14
3.3 Making data interoperable 17
3.4 Increase data re-use (through clarifying licences) 17
4 Allocation of resources 19
4.1 Estimation of costs 19
4.2 Responsibilities for data management 19
5 Data security 21
5.1 Data confidentiality and integrity 21
5.2 Data availability 21
6 iQonic and Open Research Data Pilot 23
7 Conclusion and next steps 24
8 List of Figures 25
9 List of Tables 25
10 Annex I. Data Sets 26
10.1 FRAUNHOFER 27
10.2 ATLANTIS 28
10.3 BRUNEL 31
10.4 CORE 32
10.5 SHADOW 33
10.6 HOLONIX 34
10.7 SENSAP 35
10.8 HiLASE 36

10.9 PRIMA 37
10.10 ALPES 38
10.11 FORTH 39
10.12 SACMI 40
10.13 FICONTEC 41
10.14 TAU 42
10.15 BRIGHTERWAVE 44
10.16 CRIT 45

DISCLAIMER

EXECUTIVE SUMMARY / ABSTRACT

The task T9.5, deals with Knowledge Management in which Data Management Plan (DMP) document will be delivered due M06 describing how acquired data and knowledge will be shared, made open and how it will be maintained and preserved. Identifiable project data will be provided in a manner to define relevant knowledge, increase partners' awareness, validate the result and timeframe of actions. During the proposal phase it was specified that all partners will be responsible for updating the knowledge management system periodically, with outcomes of research work conducted by other groups adding any new patents, in order to ensure ongoing relevant technological developments and innovations that are identified, analysed and built upon during the course of IQONIC project. This research will be published in confidential reports, the Plan for Exploitation and Dissemination of Results (PEDR). In particular, each partner will characterise their produced data and associated software and/or used in the project whether these are:

- Discoverable (means of an identification mechanism such as Digital Object Identifier).
- Accessible (in what modalities, scope, licenses, Intellectual Property Rights (IPR)).
- Assessable and intelligible (they allow third parties to make assessments about their reliability and the competence of the creators.
- Useable beyond the original purpose for which it was collected are usable to third parties after the collection of the data for long periods (repositories, preservation and curation).
- Interoperable to specific quality standards allow data exchange between researchers, institutions, organisations, countries, re-combinations with different datasets, data exchange, compliant with available software applications.

SCOPE

The iQonic DMP describes the observed data that are collected and processed during the life time of the project while providing the overview of available research data, access, data management and terms of use. This document is the status quo plan and determination of the partners with first assumptions of data set definitions and draft results that will be updated and improved continuously during the iQonic lifespan. As a result, iQonic will participate in the Open Research Data Pilot (ORDP) in Horizon 2020 and support the European Commission (EC) aim to improve and maximize access and reusage of research data that are generated by various projects focusing on encouraging good data management as the crucial part of research best practice.

ABBREVIATIONS

Data Management Plan	DMP
Plan for Exploitation and Dissemination of Results	PEDR
Intellectual Property Rights	IPR
Open Research Data Pilot	ORDP
Key Performance Indicators	KPI
European Commission	EC
Decision Support System	DSS
Findable, accessible, interoperable and re-usable data	FAIR
European Union	EU
General Data Protection Regulation	GDPR
European Factories of the Future Research Association	EFFRA
Uniform Resource Locator	URL
Knowledge Base System	KBS
Laser Induced Damage Threshold	LIDT
Manufacturing Execution System	MES
Enterprise Resource Planning	ERP

1 Objectives

The aim of the iQonic project is to contribute to the eradication of defects in manufacturing while providing better quality of products, increasing the flexibility and reducing the production costs. Hence, it is important for the achievements to be communicated to the public and respective targeted markets. Following the EC recommendation[1] that endorse the project results to be distributed openly to the public, iQonic will participate in the Open Research Data Pilot (ORDP) in Horizon 2020. While, the quality and performance data of the Manufacturing enterprises will be considered private and will only be available after granting permission, the research data about modelling procedures, Key Performance Indicators (KPI) validation, event modelling, inspection and real-time quality control, as well as the system optimisation, which will be collected/generated during iQonic, will be distributed freely. Therefore, a Data Management Plan (DMP) document will be developed and supported by all consortium partners (Table 1-1), which will indicate whether and how this data will be shared and/or made open and how it will be created and preserved.

Table 1-1: The iQonic project consortium

Short name	Name	Country
FRAUNHOFER (F-IOF)*	*FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.*	Germany
ATLANTIS	*ATLANTIS ENGINEERING AE*	Greece
BRUNEL	*BRUNEL UNIVERSITY LONDON*	United Kingdom
CORE	*CORE INNOVATION AND TECHNOLOGY OE*	Greece
POLIMI	*POLITECNICO DI MILANO*	Italy
SHADOW	*THE SHADOW ROBOT COMPANY LIMITED*	United Kingdom
HOLONIX	*HOLONIX SRL-SPIN OFF DEL POLITECNICO DI MILANO*	Italy
SENSAP	*SENSAP SWISS AG*	Switzerland
IP-ASCR (HiLASE)	*FYZIKALNI USTAV AV CR V.V.I*	Czech Republic
PRIMA	*PRIMA ELECTRO SPA*	Italy
ALPES	*ALPES LASERS SA*	Switzerland
FORTH	*FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS*	Greece
FILAR	*FILAR-OPTOMATERIALS SRL*	Italy
SACMI	*SACMI COOPERATIVA MECCANICI IMOLA SOCIETA COOPERATIVA*	Italy
FICONTEC	*FICONTEC SERVICE GMBH*	Germany
TAU	*TTY-SAATIO*	Finland
BRIGHTERWAVE	*BRIGHTERWAVE OY*	Finland
CRIT	*CENTRO DI RICERCA E INNOVAZIONE TECNOLOGICA SRL*	Italy

* Project Coordinator and Technical Manager within the iQonic project.

[1] Guidelines on Data Management in Horizon 2020, Version 1.0, 11 December 2013

2 Introduction

The initial data management plan is conducted and delivered due M06 of the project, that could be updated and augmented with new datasets and results during the lifespan of iQonic project. For completing the DMP deliverable, all iQonic partners have provided input following the template Horizon 2020[2] as well as the Data Set Template (Annex I). During the proposal phase, an initial DMP plan has been explored and its results are summarised and presented in the following Table 1-2. Each consortium partner will define their produced data and associated software and/or used in the project whether these are discoverable, accessible, assessable and intelligible, useable (beyond the original purpose) and interoperable. Details of the data collection purpose, data type, format, size and utility have been presented in details in sub-sections 2.1-2.5 respectfully for iQonic consortium.

Table 1-2: Research data, collected and generated during iQonic

Research Data	iQonic Partner
In-line olfactory system readings and decisions on actions	**SACMI**
Adaptive optics and in-line inspection outcomes	**FORTH**
Processed data and conclusions from the effects of Strategy(ies) activation	**ATLANTIS**
Semantic rules for Strategy activation	**ATLANTIS**
Defect detection efficiency data: false alarm rate, precision, recall, F-Measure	**ALL**
Defect prediction efficiency data: positive prediction rate	**ALL**
Thresholds for forecasting in relation to condition monitoring and defect prediction	**ALL**
Deterioration trends and intelligent engine outcomes	**CORE**
Discrete Event Modelling – cost function generation to optimize defect analysis	**BRUNEL**
Improved functionalities of i-LiKe Machine knowledge management and DSS suite	**HOLONIX**

2.1 Purpose of the data collection/generation and its relation to the objectives of the project

For F-IOF data will be used to mock up the photonics assembly chain and to use the iQonic real-time model in order to assess the photonics production and assembly procedures and try making them more efficient. By that, the iQonic model will evaluate the state of the manufacturing line and will be able to provide information (based on the data gathered) and solutions to minimize the generation of defects.

For ATLANTIS iQonic project aims at contributing to the eradication of defects in manufacturing, providing better quality of products, increased flexibility and reduced production costs. Hence, the achievements are required to be communicated to the public and the respective targeted markets. Therefore, the results are required to be distributed openly to the public and as a result iQonic will participate in the *Open Research Data Pilot in Horizon 2020*. However, the quality and performance data of the Manufacturing enterprises will be considered private and will only be available after granting permission. On the other hand, the research data about modelling procedures, KPI validation, event modelling, inspection and real-time quality control, as well as the system optimisation, which will be collected/generated during iQonic will be distributed freely.

[2] H2020 templates: Data management plan, Version 1.0, 13 October 2016

For BRUNEL data will be used by iQonic real-time model in order to assess the procedures and try make them more efficient. By that, iQonic model will evaluate the state of the manufacturing line and will be able to provide information (based on the data gathered) and solutions for us to minimize the generation of defects.

For CORE the purpose of the data collection is to use the data as inputs to machine and deep learning models for:

a) Finding anomalous operational states
b) Predicting crucial sensor values

For POLIMI data will be used for process modelling and derivation of zero-defect as well as refurbishing strategies, making use of the generated data by means of data analysis.

For SHADOW data will be used for debugging and testing robotics application.

For HOLONIX data will be stored in the KBS according to a semantic model that will simplify knowledge extraction that will be made available for the other IQONIC components to be consumed. KBS component will be also integrated in HOLONIX I-Like Machines platform to generate a set of machine/process KPIs in order to monitor the production process.

For SENSAP the collected data from the Shopfloor will be homogenised using a common data model and feed the higher-level strategies (Cognitive Functions) of the iQonic platform.

For HiLASE the purpose of data collection will be to share:

a) Capabilities of HiLASE related to LIDT measurements
b) Know-how on preparation of laser crystals and requirements for specific usage as a part of HiLASE laser systems
c) Know-how on design and production of thin-disk head

For ALPES the purpose of the data collection/generation is finding correlations between changes in the environment of production machines and mounting defects on lasers; identifying defects on laser chips.

For SACMI data will be generated by an olfactory system. This device will monitor the material used in the optoelectronic assembly to identify the presence of contaminants.

For FICONTEC data will be collected within the project to analyse the overall photonic assembly flow and test for correlations between sensor data and the quality of the final components. For this purpose, data from multiple sources like images systems, motion data, and results from alignment routines will be gathered and transferred to an external system for data evaluation. Finally, a sophisticated model for data analysis will be able to provide information and solutions to minimize the generation of defect products and thereby increasing the yield while in the same time reducing the ecological footprint.

For TAU simulation and measurement data will be generated in order to improve the performance and reliability of different active and passive components and (sub)assemblies. In-situ and post-assembly measurement data will be also collected to verify possible problems or improvements in the production chain. The collected data will be used in data analysis, iQonic project reports and deliverables, and possible journal publications.

For BRIGHTERWAVE the purpose of the data collection/generation is the improvement of the assembly yield, and the improvement of the incoming quality inspection for the components used in the assembly. The ability to use the rejected components for the other products with different characteristics.

For CRIT data will be collected and used for the communication and dissemination of the iQonic activities. Most data will be collected by partner, some could come from periodicals, scientific databases, events.

For FORTH the purpose of the data collection/generation is the correction of image aberrations.

For PRIMA the purpose of the data collection/generation is to analyse production trends; pass / fail screening; device engineering for yield improvement and failure management.

2.2 Types and formats of data the project will generate/collect

For F-IOF environmental conditions (temperature, pressure, humidity, contaminations (particles, volatile organic compounds)), components and assembly media characterization results (e.g. geometrical, optical, contamination properties, bonding agent strength) and assembly results (mainly alignment and resulting optical functionality) are the most crucial general data needed. Each use case is unique though, so different data will be needed from each case respectively. In order to evaluate which data will be needed individual use cases as well as a generalized one shall be described. The format of data is provided in XML format or excel sheets.

The origin of the data is various measurement and characterization devices (e.g. interferometers, 3D coordinate measurement machines, laser scanning microscopes, white-light interferometers, beam profilers, power meters), assembly sensors (mainly geometrical measurement systems).

For ATLANTIS the types of the data are production data, processed sensor data, raw sensor data and research data, while the formats of the data are XML, JSON, RDF. The origin of the data is SENSAP sensors for the collection of the real time data. Cameras and lasers, and maybe other measurement instruments at different points of the production line will be used.

For BRUNEL temperature, pressure, positioning are some of the general data they will need. Each use case is unique though, so different data will be needed from each case respectively. In order to evaluate which data will be needed they have to investigate each use case separately to see on what they need to focus on. The format of data which is commonly used is in XML format or excel sheets. The origin of the data is sensors, experts/machine operators, energy calculators etc.

For CORE the types and formats of data are:

a) Sensor data from production / usually JSON
b) Machine settings / database exports, CSV, XML, etc.

The origin of the data will be provided by the end-users.

For SHADOW the type and format of data will be ROS Bags http://wiki.ros.org/Bags.

For HOLONIX KBS will receive and store: raw data from sensors, structured production related data from MES/ERP systems and information acquired through UI. Messages exchanged will be in JSON format. KPIs values that will be visualized by user on the User Interface can be in graphical, table or textual format. The origin of the data is machines/systems in the Shopfloor, manufacturer's ERP/MES, User Interface.

For SENSAP the data source "types" will be of the usual data types in their raw format (e.g. INT, STRING, CHAR, REAL, etc). The set of the data source types will be identified in each use cases scenario and regarding the need of the machine learning algorithm as each use case is unique. The "kind" of the data source describes the physical property of the measured quantity with its engineering units optionally and the format will be defined by the technical partners.

The origin of the data will be from the machines (e.g. FICONTEC) and sensors (e.g. SACMI, FORTH) employed within iQonic project. In addition, data will be originated from the monitoring of the production stage/phase using the appropriate sensors that the Cognitive Functions need to operate/function (e.g. vibration sensor, temperature/humidity). Moreover, data will be generated from the vision inspection modules and through the tagging system (e.g. product location, status).

For HiLASE the type and format of the data will be in the form of reports (pdf) which may contain some drawings. The origin of the data is in-house research and development.

For ALPES the types and formats of the data are:

- Sensor data from connected fabrication equipment / JSON
- Production data and analysis data related to devices / JSON

The origin of the data is in-house generation in ALPES's production site

For SACMI the instrument generates raw data (the electrical resistances of its chemical sensors, recorded as a text file), and the recording of its output after elaboration (classification and quantification of the odour detected, recorded as a text file and as excel sheet). Diagnostic data on instrumental operation parameters (temperature, humidity, gas flow, etc.) are also recorded as text files.

The origin of the data is electrical resistance of chemo-resistors regarding odour detection, temperature and gas flow sensors regarding instrumental operation.

For FICONTEC, within the duration of the project, several different kind of data will be generated:

- Machine related data (vacuum levels, gripping forces, motor currents)
- Environmental conditions (temperature, pressure, humidity, contaminations (particles, volatile organic compounds))
- Components and assembly media characterization results (e.g. geometrical, optical, contamination properties, bonding agent strength)
- Assembly results (mainly alignment and resulting optical functionality) as final qualification parameter

Each use case is unique though, so different data will be needed from each case respectively. In order to evaluate which data will be needed individual use cases as well as a generalized one shall be described. The format of data is under discussion and will be most likely a database format.

The data provided by FICONTEC will either originate directly from the machine (motor currents, data from image analysis, vacuum and pressurized air levels) or from hardware developed by partners which will finally be installed in the assembly machine.

For TAU sensor, imaging, and simulation data will be generated and collected. Numerical data will be stored mainly in ASCII format and imaging data will be stored in typical lossy or lossless image formats.

Sensor and imaging data originates from the pre-, insitu, and post-measurements of the assemblies. Moreover, the simulation data is generated in the assembly design stage.

For BRIGHTERWAVE data format will be basic ASCII-data. The origin of the data is measured data from various sources: spectra from optical spectrum analyser, light-current-voltage characteristics for LIV measurement system.

CRIT will collect data regarding partners' organisations, staff members and activities as well as general news, events and publications. Data can be textual data or images, videos, mixed media and will be mostly in a digital format, but there could be hard copies too.

FORTH will generate/collect imaging data in typical formats (.dat, .tiff). The origin of the data is Adaptive Imaging Microscopy.

For PRIMA the types and formats of data are IR and visible images collected from cameras; data measured on laser chip. The origin of the data is PRIMA diodes production line.

2.3 Existing data, re-use and how?

Data from F-IOF as an R&D institute is only resulting from experimental investigations and prototypical development and shall only be used in exceptional circumstances (e.g. to fill a gap in the whole process chain).

For ATLANTIS it should be encouraged to make existing data available for research. All partners will be responsible for periodically updating the Knowledge management system with outcomes of research work conducted by other groups and any new patents, i.e. to ensure that ongoing relevant technological developments and innovations are identified, analysed, and hopefully built upon during the course of the project.

For BRUNEL existing data from the manufacturer and end-user are required for modelling purposes and to comprehend how the equipment and production line were working so far and get a better insight of the manufacturing processes.

CORE will use historical data to enhance the learning capability of the algorithms.

For HOLONIX existing data stored in manufacturer's system will be stored on the KBS and made available to the other partners to be consumed.

SENSAP systems will mainly stream raw data produced by sensors in the Shopfloor or transform them. SENSAP will reuse data only to evaluate the vision inspection system.

For HiLASE existing data will be used as starting point for definition of requirements regarding crystal quality (including LIDT of coatings).

For ALPES sensor data is consumed / analysed by other project partners.

SACMI will re-use data for optimization of the sensors working parameters.

FICONTEC may only use existing data to discuss data formats with partners or to provide test data for first tests of the overall process chain for data evaluation.

For TAU existing data collected from TAU's previous projects and initiatives may be used as a reference to indicate malfunctions or performance improvements. Moreover, existing data provided by other sources such as project partners may be used for the same purposes.

BRIGHTERWAVE and CRIT will not re-use any existing data.

For PRIMA data are generated continuously in production line.

2.4 Expected size of the data

For F-IOF the expected size of the data is kB to MB range.

For ATLANTIS the expected size of the data is to be evaluated during the course of the project. The expected size depends on the extend and the nature of the data that are made available. XML/JSON format data are estimated to be at least 300 MB per day.

For BRUNEL, HOLONIX, SENSAP and CRIT the expected size of the data is not yet known.

For CORE the expected size of the data is of the order of several GB.

For HiLASE the expected size of the data is not more than 100 GB.

For ALPES and PRIMA the expected size of the data is less than 1 GB.

For SACMI the expected size of the data is about 10 MB per day of measurement.

For FICONTEC the expected size of the data is kB to MB range. In case of permanent 24/7 data logging in a production atmosphere several tens of MB may be acquired.

For TAU depending on the scope of the characterization studies, the data size can vary from a few GB to a few tens of GB.

For BRIGHTERWAVE the expected size of the data can be Megabyte or probably less for each component, but the number of components in a lot can be 100's or even 1000's.

For FORTH the data size is of the order of few MB.

2.5 Data utility

For F-IOF and FICONTEC data might be useful to all project partners involved, mainly the software intelligence partners.

For ATLANTIS data might be useful to iQonic consortium and specially the end users.

For BRUNEL, SENSAP and SACMI data might be useful to all project partners involved.

For CORE data might be useful to them.

For HOLONIX data might be useful to all the project partners that perform data analytics.

For HiLASE those data will be mostly useful to FILAR.

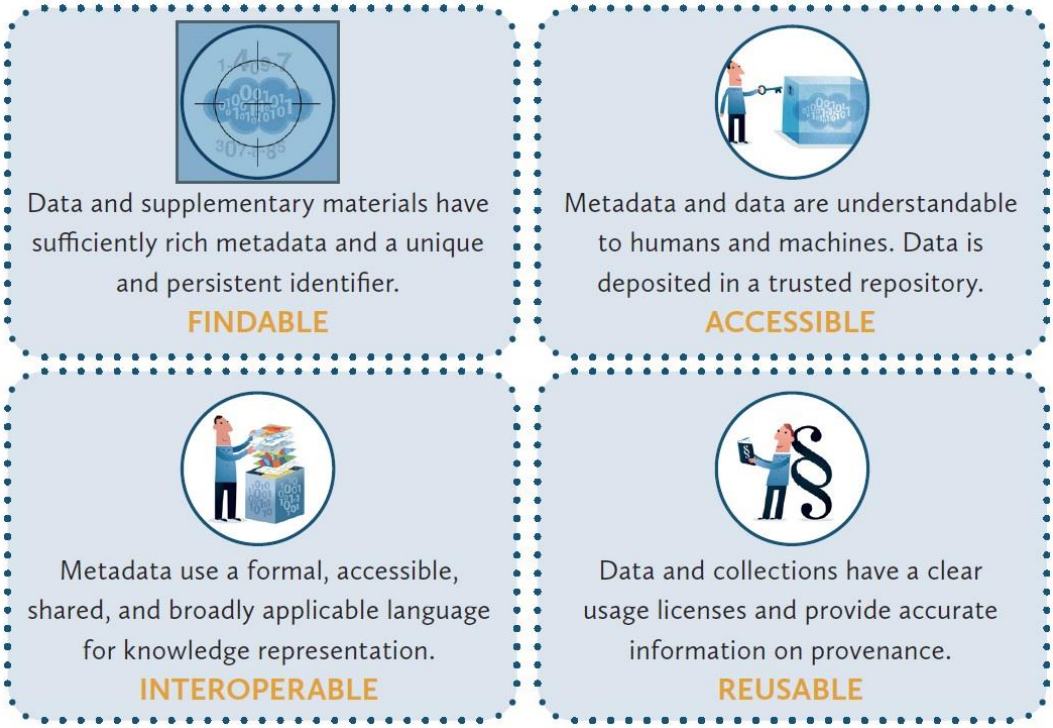
For TAU data will be useful for researchers, teachers, students, product developers and manufacturers, and other people interested in micro-manufacturing.

For CRIT data will be namely useful to project partners as well as stakeholder and potential beneficiaries.

3 FAIR data

When we are talking on research data, it is good for them to be findable, accessible, interoperable and re-usable (FAIR), presented in the Figure 1-1. The mentioned principles forego implementation choices and do not necessarily propose an explicit technology, standard or implementation solution. According to Wilkinson et al., 2016[3] the FAIR data principles are a set of guidance for management and procuration of scientific data that are relevant to all actors and participants (academia, industry, funding agencies, scholarly publishers, etc.) in the existing digital ecosystem addressing directly data publishers while promoting the maximum usage of research data. Furthermore, FAIR principles stress the importance of increasing the machine ability to automatically find and use the data while supporting individuals' reuse. In the following sections 3.1 - 3.4 iQonic consortium have addressed this matter in details.

Figure 1-1: FAIR data[4]



3.1 Making data findable, including provisions for metadata

For F-IOF, SENSAP and FICONTEC data should not be findable, since data is individual and often confidential or proprietary.

For BRUNEL results that are confidential and can only be distributed between the partners of the project won't be discoverable from third parties.

For HOLONIX in a first stage, data will be available only for members of IQONIC consortium. A possible disclosure must be agreed with the consortium member.

For SACMI data should not be findable.

3.1.1 Metadata provision

For F-IOF the created metadata is information about process parameters and results. Process parameters have a broad variability, as F-IOF covers the whole photonics components manufacturing and system assembly

[3] The FAIR Guiding Principles for scientific data management and stewardship
[4] Implementing FAIR Data Principles: The Role of Libraries

process chain. Process results most commonly are components characterization, alignment and optical performance parameters.

For ATLANTIS metadata standards that are compatible with iQonic consortium will be used. The dataset might be accompanied with a documentation of its content. Indicative metadata may include:

- Location, date, timestamp, etc., and production process that led to the defect generation.
- Annotated detection asset's vitals, asset's RULs.
- A defect detection event in the production line, the cause, the origin, the value of the defect, thresholds, the current production stage.

For BRUNEL the created metadata is information about the state of equipment, defect formation and prediction, alarms, potential correction actions, circularity index.

For SENSAP the collected data are being assigned a timestamp and location description as a metadata. In addition to that, the vision inspection module generates as a metadata of the inspected image the set-up parameters and the result (evaluation).

For HiLASE the file will always contain date and author (HiLASE) info. In the file will be clearly stated (IP) rights and technical metadata.

For SACMI metadata is diagnostic data on instrumental operation parameters. A timestamp is assigned to each measurement, for which environmental parameters (humidity, temperature) and calibration coefficients of the chemical sensors are recorded as well.

For FICONTEC the exact location and time when the data is gathered in the assembly process flow is known and shared to all partners.

For TAU metadata contains identifiers, dates, creators, formats, and the description of parameters. Furthermore, metadata includes information about how the data was collected and created. TAU follows the metadata standards that are compatible with iQonic consortium.

For BRIGHTERWAVE the created metadata will be measurement date and time, component IDs, measurement system, parameters such as wavelength range used in the measurement.

PRIMA will follow existing SQL database.

3.1.2 Standards for metadata creation

No standards for metadata creation defined yet, to be discussed within the iQonic consortium.

3.1.3 Approach towards search keyword

For F-IOF and SENSAP providing keywords is not common and depends on the set and amount of data.

For ATLANTIS specific filters, based on the metadata elements, will allow to refine the search across datasets (e.g. search dataset by specific organization, component, keywords, temporal or spatial coverage of the data etc.).

For BRUNEL the data will be processed in a format that can be searched and used electronically by other stakeholders.

For HiLASE search keywords will be provided that optimize possibilities for re-use. For example type of crystal, parameters, type of testing, type of usage, etc.

For ALPES, SACMI and PRIMA search keywords will not be provided.

For FICONTEC providing keywords is not common. As the data will be provided as database row and column headers are available and can be considered as keywords.

For TAU search keywords such as assembly parameters can be provided. The data stored in EUDAT services can be searched using B2FIND data finder.

For BRIGHTERWAVE and CRIT it is not defined yet whether search keywords will be provided that optimize possibilities for re-use.

3.1.4 Use of persistent and unique identifiers

For ATLANTIS Digital Object Identifiers is a standard identification mechanism that could be used.

For HiLASE some data will be part of scientific articles with Digital Object Identifiers.

All devices produced at ALPES have a unique identifier. Data related to devices will clearly be linked to this identifier which will be anonymized, if this appears suitable.

For TAU components and subassemblies can be registered with keywords, component and batch identifiers. Hence, the data, which is related to the components and subassemblies, can be searched afterwards based on attached metadata from the used storage database.

For PRIMA all devices data produced have a unique identifier, being time stamp or production line related information.

3.1.5 Naming conventions

For F-IOF there is no standard yet, but there is an obligation in the FRAUNHOFER society that all data (filenames) end with either "_pub" (public), "_conf" (confidential), or "_int" (for internal use only).

For ATLANTIS for metadata, dataset and template names a naming convention could consist of :

- A prefix, indicating if it is a dataset, a metadata or a template and a short and meaningful name of the dataset/template.
- A root composed by the name of the data provider organization.
- A suffix indicating the name of the component which will used by the organization.

For HiLASE an example of naming convention is 20190212_IQONIC_HiLASE specification_v1-01.docx.

SACMI does not follow naming convention. Output file names are user-defined.

For SENSAP, ALPES, FICONTEC, BRIGHTERWAVE and CRIT there is no standard yet, to be discussed within the consortium.

For TAU data is specified according to lot number or batch identifier or assembly id or component identifier. The lot number will provide a unique identifier for fabricated assembly process type. Batch identifier provides name for the fabricated batch for each assembly process type. The assembly id specifies each assembly and the component id specifies the components in the assembly.

For PRIMA there is no standard defined yet, but already existing complete SQL database is set, so worth keeping existing conventions.

3.1.6 Clear versioning

F-IOF, BRUNEL, CORE, HOLONIX, SENSAP, HiLASE, ALPES, SACMI, FICONTEC, BRIGHTERWAVE and PRIMA provide clear version numbers.

For ATLANTIS the versioning management of the data, metadata template and in general the files stored into the Repository could be applied at two levels:

1. Via the naming convention and the use of the date as suffix, indicating the last version of the file uploaded into the Repository.
2. Use a Repository which can support a simple version control system for the uploaded files.

For TAU different versions are specified based on the dates.

3.2 Making data openly accessible:

For F-IOF openly available as the default will be results of prototypical processes that F-IOF implemented for the iQonic project on its assembly machines, if required e.g. for filling gaps in the assembly chain. F-IOF, BRUNEL and SENSAP believe that it should be considered for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out. As the project evolves the depth and breadth of such information will be ascertained.

For F-IOF, SENSAP, SACMI and FICONTEC a dedicated data repository for the project is provided by the FRAUNHOFER Owncloud service. At the moment the service is used to store project info there. If it is suitable

also as a sharing platform for process data is tbc. At the moment the FRAUNHOFER Owncloud is defined to be the data sharing platform, it is tbc if another type of deposition platform with more options for access is required.

For F-IOF documentation about the software needed to access the data will be included, but in most cases it will not be possible to include the relevant software, as proprietary software is used to create most of the process data. For F-IOF, SENSAP, SACMI and FICONTEC if there are restrictions on use, the Owncloud is accessible via username and password privileges for individuals from each iQonic partner. The identity of the person accessing the data will be ascertained by username and password.

For ATLANTIS the research data about modelling procedures, KPI validation, event modelling, inspection and real-time quality control, as well as the system optimization, which will be collected/generated during iQonic, will be distributed freely. The achievements are required to be communicated openly to the public and the respective targeted markets. The results are required to be distributed openly to the public. On the other hand, the quality and performance data of the Manufacturing enterprises will be considered private and will only be available after granting permission.

Relational databases or other types of databases can be used to publish data. Database management software tools are needed to access the data. The data and associated metadata, documentation and code will be deposited to certified repositories which support open access where possible.

Related manuals about the software needed to access the data will be provided and open source code could be stored to GitHub. If there are restrictions on use, access will be provided with different access level for each different user group. The identity of the person accessing the data will be ascertained with the usual authentication methods and the use of different credential per user. Appropriate access levels will be granted by the creation of groups so that there will be protection against unauthorised access.

For BRUNEL openly available will be the results of the modelling (e.g. Energy consumption, Process Time etc.). Datasets which cannot be shared due to confidentiality won't be openly available (e.g. methods of process, trivial information on the production line of an industry etc.).

The data will be made accessible through direct access to the application software and the stored data can be stored either locally or/and in the dedicated consortium cloud services. What methods or software tools are needed to access the data depends on the protocols and solutions provided by their service integration partners. BRUNEL will work with those partners to facilitate the process. Where the data and associated metadata, documentation and code will be deposited, will be determined by system integration and management partners.

Documentation about the software needed to access the data will potentially be included. Whether it will be possible to include the relevant software will be determined by the nature and requirements of the end user and proprietary applications in existence. If there are restrictions on use, access will be provided through username and password privileges. The identity of the person accessing the data will be ascertained by username and password.

For CORE which data will be openly available as the default will be at the discretion of the end-users. Usually this kind of data are restricted from publication. Data will be accessible to CORE through HOLONIX's iLike platform. Usually python requests library, if provided via REST API will be needed to access the data.

CORE will not share end-users' data. It will be used internally for development. Documentation is found in the GitHub repository. The relevant software is already open source and easily found. CORE algorithms are proprietary.

For SHADOW data will be made accessible via GitHub repository. Web Browser and ROS bag viewer http://wiki.ros.org/rqt_bag are needed to access the data. Documentation about the software needed to access the data is included: http://wiki.ros.org/rqt_bag. ROS bag viewer is open source https://github.com/ros-visualization/rqt_bag.

For HOLONIX data will be shared outside IQONIC consortium only if explicitly agreed by consortium. HTTP request tools are needed to access the data. The data will be made accessible via APIs. Documentation about the software needed to access the data will be included, but it is not possible to include the relevant software. If there are restrictions on use, access will be provided through authentication and authorisation mechanisms implemented upon available APIs.

For SENSAP as the collected data originates from machines and sensors, in a production Shopfloor from consortium partners customers, the data will follow their model regarding openness to the public. An ftp client supporting ssh is needed to access the data. Documentation about the software needed to access the data will be included and it should be possible to include the relevant software.

For HiLASE data on LIDT testing are datasets, which will be up to FILAR to decide level of accessibility to other partners. HiLASE will keep right to preserve details on laser design used for testing crystals. Design of lasers is IP of HiLASE. HiLASE will also restrict sharing of data on thin disk design and bonding procedures, while this is vital IP.

The data produced during the project can be part of some articles and as such they will be made accessible. The data will be provided in the pdf or docs format. Ordinary office software will be enough to access the data. Documentation about the software needed to access the data will be included. The data can be part of open access journals relevant for field of laser development and design. Elsevier as publisher of journals in the field of Physical and Sciences and Engineering propose to use Data in Brief and open access approach. Other option will be to use Mendeley Data. Mendeley Data is open and free of charge and some journals allows respiratory of submitted manuscript.

For ALPES there are no restrictions on sensor data generated on production machines / environmental data; ALPES does not share publicly data directly related to devices / products they fabricate. Data related to ALPES products and fabrication process is strictly confidential. Data for the project will be provided in open formats. Sensor data acquired in ALPES's facilities will be provided to project partners consuming the data and analysing it together with associated device-relevant data; ALPES will not store that data. Documentation about the software needed to access the data will not be included.

For SACMI data directly generated by the olfactory system will be available to involved project partners, but not available outside the project. No specific software is required, data are stored as text files. To access Excel sheets, Microsoft Office or OpenOffice are needed. Documentation about the software needed to access the data will not be included. The relevant software will not be included, but many open source options are available.

For FICONTEC as all machine related data originates from customer components and can be traced back to those, all datasets should be considered as confidential per default. After discussion with the end users and if agreed some not component relevant data could also be made public. Depending on the confidentiality of the data and the size also an distribution via email is possible.

The database can be accesses with different tools. Depending on the size of the data a direct reading in in the following software tools should be favoured. Documentation about the software needed to access the data should be included.

For TAU any data that may be defined as confidential or private according to the consortium agreement is not disclosed. This confidential or private data may include trade secrets, patents, or personal information. If data is defined as public according to the consortium agreement, it can be made openly accessible by default.

For BRIGHTERWAVE by default, everything is confidential. They have their own new disruptive laser technology, which is targeting biomedical imaging as well as consumer electronics applications. If successful, it is extremely likely that their competitors will try to copy the technology. Openly available data will make their task a lot easier. How the data will be made accessible is to be decided. They have their own server for which data is stored. The part of the data can be shared and accessed within the consortium. Basic software is enough to access the data. If more complicated software is needed at the later stage it will be included.

For CRIT data resulted from the communication strategy will be available for everyone to see; while data for dissemination purposes will be open only after non-disclosure partners' check. Data will be published on the project's website and social network pages, however they will also be stored in CRIT's databases.

For PRIMA only a sub-set of data will be shared to keep proprietary PRIMA solutions and sensible PRIMA data. How will the data be made accessible is to be agreed with PRIMA ICT and it should be compliant to PRIMA security. Data are CSV, excel files and ASCII files generated from SQL database tables.

3.3 Making data interoperable

For F-IOF, SENSAP and FICONTEC data produced in the project are typically not interoperable. Tbc by the consortium, if generalized data can be created that is interoperable. For F-IOF and SENSAP it will be necessary, if not mandatory to use standard vocabularies for all data types present in their data set, to allow interdisciplinary interoperability. In case it is unavoidable that they use uncommon or generate project specific ontologies or vocabularies, they will provide mappings to more commonly used ontologies by integrating a glossary.

For ATLANTIS emphasis will be given so that produced data in the project are interoperable. It should be mentioned that there are no widely accepted standards in the market for the communication among Shopfloor h/w and s/w assets and metrology sensors and inspections systems, especially with respect to the data being exchanged and processes. iQonic will be built with architecture interoperability from the outset, using the concept of IoT middleware for interfacing with any metrology/inspection and other tools. There will be semantic annotation of the exchanged data and use of appropriate ontology elements to make data interoperable.

For BRUNEL data will be designed in a form that is platform independent as possible. However, this will only be ascertained once we establish and learn about the current state of affairs in their end-user facilities. They will use standard engineering vocabularies as much as they can. In some cases, though, they might need to generate some specific ones. They will most likely use standard vocabularies for all data types present in their data set, to allow interdisciplinary interoperability. In case it is unavoidable that they use uncommon or generate project specific ontologies or vocabularies, they will provide mappings to more commonly used ontologies.

For CORE, ALPES, SACMI and PRIMA the data produced in the project should not be interoperable.

For HOLONIX semantic framework used in the project aims to favour interoperability. SWRL (Semantic Web Rule Language) will be used to make data interoperable. A semantic will be created following standard as much as possible. In case it is unavoidable that they use uncommon or generate project specific ontologies or vocabularies, they will provide mappings to more commonly used ontologies.

For HiLASE data produced in the project should be interoperable. Data will be, as it was mentioned, part of publications and will be presented during certain specific conferences. They will use standard vocabularies for all data types present in their data set, to allow interdisciplinary interoperability. In case it is unavoidable that they use uncommon or generate project specific ontologies or vocabularies, they will provide mappings to more commonly used ontologies.

For TAU the shared (meta)data is made interoperable by enabling easy access to the data through databases and using (meta)data standards and vocabularies. (Meta)data standards will make the accessed data readable and re-usable.

For BRIGHTERWAVE such distribution is not desirable.

3.4 Increase data re-use (through clarifying licences)

For F-IOF, BRUNEL, SENSAP and FICONTEC we should refer to consortium agreement about how will the data be licensed to permit the widest re-use possible. Furthermore, background IP may be in use by other parties, but the foreground IP requires consortium approval based on the agreement drawn at the beginning of the project. See regulations in the CA and GA.

For F-IOF, BRUNEL, SENSAP and FICONTEC when the data will be made available for re-use should depend on the nature and direction of the developments. They will fall back on the original agreement and in cases of new scenarios, new/amendments will be made. For F-IOF, SENSAP and FICONTEC data shall be created according to good scientific practise (GSP), while for BRUNEL data quality assurance processes will be described during the research and development phase.

For ATLANTIS data will be made available for re-use. It is encouraged to make existing data available for research. It is foreseen that in six months before the end of the project, partners will be in a position to apply for any relevant patents or decide on the type of licence for the early use and exploitation of technologies.

There will be development of an acceptable exploitation agreement for the partners and for the exploitation and dissemination of the results to outside parties. Copyright statements will protect any written material produced during the project. All information supplied by any of the partners will be kept secret for a period of five years unless there are agreements between the partners via the PSC, otherwise all patents and all other publication will require prior agreement from the PSC in respect to content and the publication media. To this end, each partner should notify the consortium about the content and material the wish to publish/disseminate, and a 30-day evaluation period will be provided. This will allow the PSC to assess whether a dissemination action compromises their IPR (either background or the results) or lacks sufficient acknowledgements, materials, and other publishable knowledge.

It is intended data, which is not restricted, to be available after the end of the project for research. The PEDR will include an exploitation strategy and plan, clearly describing how exploitation opportunities will be identified and progressed both during and after the project. The consortium members will be encouraged to continue to maintain contact and work together after the project completion in order to maximize the business opportunities that will inevitably arise following the successful technology developments resulting from the project.

Data produced by CORE will be restricted to the consortium partners. How long is it intended that the data remains re-usable depends on the end-user.

For HOLONIX how will the data be licensed to permit the widest re-use possible should be agreed within the consortium. Besides, whether the data produced and/or used in the project are useable by third parties, in particular after the end of the project, should be agreed within the consortium.

HiLASE will permit the widest re-use possible by publishing data, presenting data on conferences, using open access. Typical time window for publishing (6 months) and for patenting (one year) will be applied. Some data can be particularly protected if the data are of high value to HiLASE and its further development. In case of important and patented data results, they will remain re-usable for few years.

For ALPES data produced and/or used in the project should not be useable by third parties, in particular after the end of the project.

For SACMI how will the data be licensed to permit the widest re-use possible and when will the data be made available for re-use should be regulated by consortium agreement. Whether the data produced and/or used in the project should be useable by third parties, in particular after the end of the project, should also be regulated by consortium agreement and grant agreement. How long is it intended that the data remains re-usable is not yet defined.

For TAU (meta)data are published with a broadly accessible data usage licence. Furthermore, the provenance of the data is provided in the metadata. The data usage license follows the commonly agreed guidelines of iQonic consortium. The generated data will be made available for re-use after the project. The used components, assembly stations, or process chains can contain trade secrets, patents, and intellectual rights that may restrict the use of data. Otherwise, TAU will enable the re-use of data after the project end for third parties. TAU will follow the convention of other project partners about how long it is intended that the data remains re-usable.

Data quality control is used to ensure that no data is accidentally changed during the technical handling, converting and transferring data, or processing and analysis. Data quality control is also used to confirm that the accuracy of the data is maintained over its entire life cycle. The data quality control includes conversions from analog to digital signals as high resolution as possible. Moreover, original information content is also stored in all conversions.

For BRIGHTERWAVE data licensing is not on their plans and data is not planned to be re-used. The figures of merit such as obtained improvement in the assembly yield can be freely used in publications etc. The data such as output power variation as a function of the distance between the optical elements is not made public, since this reveals too much info on the confidential technology.

For PRIMA any data re-use is not possible.

4 Allocation of resources

Further to the FAIR principles, DMP also address the allocation of resources, estimating the costs for making the data FAIR while identifying the responsibilities for data management in project describing the costs and potential value of long-term preservation. The consortium has provided detailed information summarised in sections 4.1 and 4.2, following the Horizon 2020 guideline and specific questions on FAIR Data Management[5]:

- What are the costs for making data FAIR in iQonic project?
- *How will these costs be covered?
- Who will be responsible for data management in iQonic project?
- Have the resources for long term preservation been discussed, who decides and how, what data will be kept and for how long they will be kept?

*Note: The costs related to open access to research data are eligible as part of the Horizon 2020 grant if compliant with the Grant Agreement conditions.

4.1 Estimation of costs

For F-IOF and SENSAP the costs for making data FAIR in iQonic project are to be confirmed. How the costs will be covered is to be confirmed.

For ATLANTIS the costs for making data FAIR in iQonic project cannot be figured out yet. It is possible that costs related to open access to research data to be covered by Horizon 2020 grant if compliant with the Grant Agreement conditions.

For BRUNEL their administrators may be able to provide information about the costs for making data FAIR in iQonic project when and where required. How the costs will be covered is to be explored.

For HOLONIX the costs for making data FAIR in iQonic project are to be defined. How the costs will be covered is to be defined.

For HiLASE the costs for making data FAIR in iQonic project are for Open Access publishing. The costs will be covered by using Open Access which can be covered by Horizon 2020 grant.

For SACMI the costs for making data FAIR in iQonic project are not defined yet, they will be if necessary. How the costs will be covered is not defined yet.

For FICONTEC the costs should not be high. However, it has to be decided and agreed on by all partners from whom data will be collected. The costs will mainly be personal costs and hence can be charged to the project in the normal way.

For TAU the costs for making data FAIR in iQonic project are personnel costs and quota costs. The costs, that will come from storing the data, are covered by the H2020 grant.

For BRIGHTERWAVE the costs for making data FAIR in iQonic project have not been investigated. How the costs will be covered is not known.

For CRIT the costs for making data FAIR in iQonic project are not known yet. The costs will be covered from communication budget.

4.2 Responsibilities for data management

ATLANTIS will lead the task for data management with consultation and support from all partners. The resources for long term preservation have not been established yet.

For F-IOF, BRUNEL, SENSAP, SACMI and FICONTEC the consortium dedicated working group will be responsible for data management. At F-IOF the project manager (and overall coordinator) Erik Beckert is responsible for data management. At SENSAP the project manager (and overall coordinator) Andreas Mantelos

[5] Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26 July 2016

is responsible for data management. At FICONTEC the project manager (and overall coordinator) Moritz Seyfried is responsible for data management.

For F-IOF, SENSAP and FICONTEC the resources for long term preservation are to be confirmed.

For BRUNEL and SACMI the resources for long term preservation have not been discussed yet.

For CORE the resources for long term preservation are to be defined.

For HOLONIX ATLANTIS will be responsible for data management in iQonic project. The resources for long term preservation have not been discussed yet.

For HiLASE scientists involved in the project with help from IT support team will be responsible for data management. HiLASE has its own server which can be used for long term preservation of data.

For ALPES lifecycle of device-related data is decided on a case-by-case basis, depending on client, pertinence of data.

For TAU Heikki Virtanen will be responsible for data management. The project responsible person determines the value of the data. The project data management leader of TAU determines the storing costs and decides how long the data will be stored after the project. In this task, the terms and conditions of the consortium agreement and the data management instructions of TAU are followed.

For BRIGHTERWAVE Dr. Antti Laakso and Dr. Janne Konttinen will be responsible for data management. The resources for long term preservation have not been discussed.

For CRIT Margherita Animini will be responsible for data management. The resources for long term preservation are not known yet.

For PRIMA its Production Department will be responsible for data management.

5 Data security

The iQonic consortium will follow a strict policy regarding the privacy and data protection as a fundamental principle and hence apply and comply with the European Union (EU) 2016/679 General Data Protection Regulation (GDPR)[6] on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Apart from respecting this legislation, the iQonic consortium partners will respect the procedures implemented for data collection, storage, access, sharing policies, protection, retention and destruction according to the requirements of the national legislation and EU standards[7]. The details and information based on following questions below, have been provided by all iQonic partners, summarised and presented in sections 5.1 and 5.2 respectfully.

- What provisions (including data recovery, secure storage and transfer of sensitive data) are in place for data security?
- Is the data safely stored in certified repositories for long term preservation and curation?

5.1 Data confidentiality and integrity

For F-IOF, SENSAP and FICONTEC https shall be used for transfer; standard 128-bit AES WinZip encryption shall be used for the storage of data (tbc).

For ATLANTIS data is going to be persisted to relational database system. A regular back up service will run in the background. Maybe encryption of sensitive data will be used.

For BRUNEL encryption will be used throughout the whole transfer and storage of data.

For CORE data are used offline in the model training phase. The data generated by the inference engines will be made available to the end-users through authentication APIs.

For HOLONIX data will be stored in secure servers owned by HOLONIX for the duration required by iQonic project.

HiLASE has IT support team, which is taking care of their own server, back up data and implementation of data security procedures.

ALPES does not store sensor data produced in their facility; business-related data is stored in-house.

For SACMI there will be restricted access to data recorded by the olfactory system.

For TAU periodical database snapshots from the stored data will be taken and integrity checks will be carried out in the data storages. The data will be also protected against viruses and other similar threats.

BRIGHTERWAVE has a-state-of-the-art extremely secure IT system in use.

CRIT will put into place provisions already in use internally that are also compliant with the GDPR provisions.

PRIMA is providing internally to data security. Only a sub-set of data will be provided to the consortium members.

5.2 Data availability

For F-IOF, ATLANTIS, BRUNEL, HiLASE, SACMI, FICONTEC, BRIGHTERWAVE and CRIT the data will be safely stored in certified repositories for long term preservation and curation.

For F-IOF, SENSAP and FICONTEC confidential and proprietary data can have an impact on data sharing. No personal data shall be collected. For data management some data storage procedures provided by German government are mandatory for F-IOF, while for SENSAP and FICONTEC no use of other national/funder/sectorial/departmental procedures is made for data management.

[6] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016
[7] Ethics and data protection, 14 November 2018

For ATLANTIS if there are any ethical or legal issues that can have an impact on data sharing they could be discussed in the context of the ethics review. There will be informed consent for data sharing and long term preservation in case of questionnaires dealing with personal data. For the moment no use of other national/ funder/sectorial/departmental procedures is made for data management.

For BRUNEL there might be confidentiality issues (legal issues) that can have an impact on data sharing.

For CORE and ALPES data should not be stored in certified repositories for long term preservation and curation. Also there are no ethical or legal issues that can have an impact on data sharing. For ALPES no use of other national/funder/sectorial/departmental procedures is made for data management.

For HOLONIX data will be stored in secure servers owned by HOLONIX for long term preservation and curation. Confidentiality related to machines/process related data can have an impact on data sharing.

For HiLASE they are not aware for the moment for any ethical or legal issues that can have an impact on data sharing. Also for the moment no use of other national/funder/sectorial/departmental procedures is made for data management.

For SACMI confidential data and intellectual property can have an impact on data sharing. No personal data will be collected. No use of other national/funder/sectorial/departmental procedures is made for data management.

For TAU EUDAT services should be able to provide the long term preservation and curation of data. TUT will obey the terms and conditions of the consortium agreement, patents, intellectual rights, the data sharing instructions of TUT, and Finnish laws. TUT is not aware of any other ethical or legal issues that have impact on data sharing. Personal data is not collected. TAU will follow an Open Science and Research policy and the national guidelines of data management.

For BRIGHTERWAVE and PRIMA there are no ethical or legal issues that can have an impact on data sharing. Also no use of other national/funder/sectorial/departmental procedures is made for data management.

For CRIT there are some ethical and legal issues that could arise in case of data sharing and they are addressed in the D1.1 and D1.2. There will be informed consent for data sharing and long term preservation in case of questionnaires dealing with personal data. For the moment no use of other national/funder/sectorial /departmental procedures is made for data management.

For PRIMA data is safely stored internally to PRIMA Diode Fab for long term preservation and curation.

6 iQonic and Open Research Data Pilot

Open Research Data Pilot was launched by the EC, along with the Horizon 2020 programme, whose purposes was to improve and take full advantage of the research data produced by various EU projects. The active approach to research data management provides the benefit of fast and easy access, efficiency, advanced quality and visibility of research. In this context, certain data that are produced during the lifetime of the iQonic project will be published with open access respecting the objective of IPR and data privacy principles. Each consortium partner will define the produced data during the lifetime of the project, specifying whether they are discoverable, accessible, assessable, intelligible, useable beyond the original purpose and interoperable. Several applications will be developed and tested in real industrial environments that will led to creation of a significant number of different datasets. In order to collect the information about the research data that will be generated in iQonic project, from different project activities, we have summarised the partners assumptions and possible expectations of data sets in Annex I section presenting the details:

- Data set reference and Name: Name, Homepage, Publisher, and Maintainer.
- Data Set description: Description, Provenance, Usefulness, Similar Data, Re-Use and integration.
- Standards and Metadata: Metadata description, Vocabularies and Ontologies.
- Data Sharing: License, URL Data Set Description, Openness, Software Necessary, Repository.
- Archiving and preservation: Preservation, Growth, Archive, Size.

Such data can be anonymised for statistical or other purposes and shared with open access, which could be further analysed and provide the possibility to extract information and knowledge from them. Each dataset can be accompanied by several metadata (e.g. type, gender, age, etc.) which can support various kinds of historical data analysis.

One of the possible ways to support DMP could be through the Research Data Alliance that provides a Metadata Standards Directory whose goal is to develop a collaborative, open directory of metadata standards applicable to scientific data that can help addressing the infrastructure challenges. Also, the EUDAT B2SHARE tool can be considered as the DMP support as well, since it is familiar that the tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

To ensure visibility and openness of iQonic resources, some platforms have been considered to be used, where general public, researchers and other investigators can discover and download information, data and documents on the project's results. The platforms and widely used research data repositories that are considered for iQonic, which allow research stakeholders to search and retrieve vilely open and all types of data that are uploaded by other researchers, are zenodo and EC's OpenAIRE platform. Furthermore, the EFFRA Innovation Portal is also taken into consideration since being provided by the European Factories of the Future Research Association (EFFRA), is a unique resource database combining various projects' databases' information about the latest projects' outputs, together with reports and demo materials.

7 Conclusion and next steps

The initial DMP is presented and delivered in this document due M06 describing how acquired data and knowledge will be shared and/or made open as well as how data will be maintained and preserved during and after the timeline of the iQonic project. Each partner has provided answers within the H2020 questioner survey according to the provisions of the national legislation that are in line with the respective EU Directives for Data Management and ethics. Project data will be identified and they will be provided in a way to define relevant knowledge, increase partners' awareness, validate results and timeframe of various actions. Additionally, all partners will be responsible to periodically update knowledge management systems, databases and platforms described in section 6 providing datasets that can be used for statistical and/or other purposes and shared as open access respecting the objective of IPR and data privacy principles. Further on, the outcomes of conducted research work and ongoing relevant technological developments and innovations that will be identified, analysed and built upon during the course of project, will be published in the confidential reports PEDRs.

8 List of Figures

Figure 1-1: FAIR data 12

9 List of Tables

Table 1-1: The iQonic project consortium 6
Table 1-2: Research data, collected and generated during iQonic 7

10 Annex I. Data Sets

The data set template was created by ATLANTIS in order to collect and summarise the partners data set, activities and responsibilities, while providing information about metadata and how, where and for how long the data will be stored including the backup (respective sub-sections below). The iQonic partners support the project being open and each partner provided the assumptions based on their overall experience, as data formats and data handling procedures in this stage of the project are not yet fixed.

10.1 FRAUNHOFER

Data Identification	
Data set description	As an R&D institute, F-IOF does not create industrial data sets. Nevertheless, during prototypical process developments R&D data sets are created that relate to individual process steps along the process chain of photonics assembly. Typical sets of data comprise information about process parameters and related processing results, but no generalized description can be given.
Source (e.g. which device?)	Data sources are devices that are used during the whole process chain: • Cleaning (e.g. ultra-sonic bath chemics and temperature parameters). • Preparation (e.g. adhesion tests from witness samples of coating processes). • Alignment (data of active and passive alignment measurement devices). • Attachment (pre- and post-bond alignment state, bonding strength). • Characterization (tactile, optical profilometry and 3D coordinate measurement to characterize components and systems).
Partners activities & responsibilities	
Partner owner of the device	F-IOF.
Partner in charge of the data collection (if different)	N/A.
Partner in charge of the data analysis (if different)	BRUNEL, CORE, ATLANTIS, HOLONIX.
Partner in charge of the data storage (if different)	TBD with technical partners.
WPs and tasks	WP3/T3.4 and WP4/T4.1.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	• Each produced data can be attached with a timestamp. • Each produced data can be attached with a parameter stamp. • TBD with technical partners: unified parameter.
Standards, Format, Estimated volume of data.	The format of the data can follow the common data model (WP2, WP3).
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	F-IOF's data shall allow to cover missing links in the process chain, as the other consortium members might not be able to provide data along the full process chain.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The data is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Data sharing within the consortium is allowed. Data sharing and re-distribution outside the consortium requires confirmation by F-IOF, as data might comprise proprietary information.
Embargo periods (if any)	TBC.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	TBC.

10.2 ATLANTIS

Data Identification	
Data set description	The real time data is collected by the SENSAP sensors. Data will be used to analyse defect types and analyse common causes of defect and produce classifications in terms of defect severity, occurrence and impact on involved process chain. Recommendations will be provided to DSS combining product monitoring models and data analytics from heterogeneous sources of monitoring-diagnosis-prediction-detection. DSS will store and analyse the effects and impacts from Strategies execution.
Source (e.g. which device?)	Cameras, lasers and measurement instruments at different points of the production lines.
Partners activities & responsibilities	
Partner owner of the device	The device will be owned to the industry, where the data collection is going to be performed.
Partner in charge of the data collection (if different)	SENSAP (maybe also other various partners related to the specific incident and/or operation).
Partner in charge of the data analysis (if different)	ATLANTIS will perform data analysis related to recommendations for DSS. Various partners related to the specific incident and/or operation (maybe POLIMI, CORE, BRUNEL can also contribute).
Partner in charge of the data storage (if different)	ATLANTIS will persist the data related to DSS into a relational database. Various partners can handle the rest of the data (maybe POLIMI, CORE, BRUNEL can also contribute).
WPs and tasks	WP5/T5.2.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	The dataset might be accompanied with a documentation of its content. Indicative metadata may include: • location, date, timestamp, etc., and production process that led to the defect generation. • annotated detection asset's vitals, asset's RULs.
Standards, Format, Estimated volume of data.	The data will be stored at XML/JSON format and are estimated to be at least 300MB per day.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The DSS will use the data to optimize response to defects. DSS is related to the optimization of process chains operations vs. defect management. DSS will incorporate autonomous and hierarchical decision support with respect to provided recommendations. The DSS will analyse the effects and impacts from Strategy execution, through its machine-learning element, in order to improve efficiency of future recommendations, both for process-chain optimization and for Strategies activation.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The full dataset will be confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Most suitable protocols for data sharing are RabbitMQ and HTTP Restful webservices but data sharing is yet to be decided along with the industrial partners.
Embargo periods (if any)	None.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data is going to persisted to relational database system. A regular back up service will run in the background to ensure data reliability and to avoid data losses. Optionally, an aging algorithm may be used to decide which records are too old and need to be removed.

Data Identification	
Data set description	The real time data is collected by the SENSAP sensors. Data will be used to define the severity of the defect and decide future actions. The generated suggestions by the RSC component can be sent as notifications (SMS or email) to the addressed operators, to execute the appropriate actions.
Source (e.g. which device?)	Cameras and lasers (maybe other measurement instruments at different points of the production line).
Partner sactivities & responsibilities	
Partner owner of the device	The device will be owned to the industry, where the data collection is going to be performed.
Partner in charge of the data collection (if different)	SENSAP (maybe also other various partners related to the specific incident and/or operation).
Partner in charge of the data analysis (if different)	ATLANTIS.
Partner in charge of the data storage (if different)	• ATLANTIS will persist the data into a relational database. • HOLONIX will utilize that data to build the RDF data.
WPs and tasks	WP5/T5.5.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	The dataset might be accompanied with a documentation. Possible metadata include: • location, date, etc., and production process that led to the defect generation, • a defect detection event in the production line, the cause, the origin, the value of the defect, thresholds, the current production stage.
Standards, Format, Estimated volume of data.	The data will be stored at JSON format and are estimated to be at least 300MB per day.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The RSC will use the data to decide whether or not a defective part/product should return to a previous production stage.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Most suitable protocols for data sharing are RabbitMQ and HTTP Restful webservices but data sharing is yet to be decided along with the industrial partners.
Embargo periods (if any)	None.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data is going to persisted to relational database system. A regular back up service will run in the background, and aging algorithm will decide which records are too old and need to be removed.

Data Identification	
Data set description	Higher level communication Middleware exchanges data with MES and other higher level management systems, like ERPS. The Middleware communicates with DSS and RSC components. Furthermore, the exchange data are going to be semantic annotated, by using the appropriate ontology based on the project's requirements. The intended communication protocols are going to be HTTP and OPC-UA.
Source (e.g. which device?)	Production data from MES, other higher-level managements systems, DSS and RSC.
Partners activities & responsibilities	
Partner owner of the device	The devices are owned by the addressed end users.
Partner in charge of the data collection (if different)	ATLANTIS (maybe also other various partners related to the specific incident and/or operation).
Partner in charge of the data analysis (if different)	ATLANTIS will perform data analysis and semantic annotation of the exchanged data of the Middleware component (maybe SENSAP, HOLONIX, and FICONTEC could also contribute).
Partner in charge of the data storage (if different)	ATLANTIS' Middleware and HOLONIX's KBS will persist the exchanged data.
WPs and tasks	WP6/T6.4
Standards	
Info about metadata (Production and storage dates, places) and documentation?	TBD.
Standards, Format, Estimated volume of data.	XML format, JSON format, RDF format
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The data will be used for better communication and monitoring of the system, optimization of the process-chain and defect management.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The full dataset will be confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Data sharing is yet to be decided along with the industrial partners.
Embargo periods (if any)	None
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data is going to be persisted to a database. In the future to define is the DB systems is a relational or NoSQL one. A regular back up service will run in the background to ensure data reliability and to avoid data losses.

10.3 BRUNEL

Data Identification	
Data set description	Real-time data will be collected by sensors installed on the manufacturing line. Historical data which might be needed will be provided by experts/machine operators etc. These data will be used for our model to predict the outcome and minimize the generation of defects.
Source (e.g. which device?)	Temperature, Energy and Pressure sensors. Also positioning (angle of machine etc.) measurements.
Partners activities & responsibilities	
Partner owner of the device	The owner of the device which will most probably be a software application containing the necessary algorithms will be embedded in the equipment or be situated alongside of the system (monitoring and alarm system). The data collection will happen at the end-user sites remotely or in person.
Partner in charge of the data collection (if different)	SENSAP (maybe also other various partners related to the specific incident and/or operation).
Partner in charge of the data analysis (if different)	BRUNEL.
Partner in charge of the data storage (if different)	HOLONIX.
WPs and tasks	WP5/T5.2/T5.3
Standards	
Info about metadata (Production and storage dates, places) and documentation?	N/A.
Standards, Format, Estimated volume of data.	Not yet known.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The model will use the data in order to assess whether a product will come out with defects or not. Then it will decide about the continuation of the procedure.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Data sharing is capable through HTTP RESTful web services and RabbitMQ Pub/Sub pattern.
Embargo periods (if any)	None
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data is going to persisted to relational database system. A regular back up service will run in the background, and aging algorithm will decide which records are to old and need to be removed.

10.4 CORE

Data Identification	
Data set description	The real time data by the SENSAP sensors along with the historical data from the end-users. Data will be used to feed the inference systems developed by CORE INNOVATION. The generated outputs will be possible anomalies and prediction of future sensor values.
Source (e.g. which device?)	Raw data from various sensors installed along the production line; process- and device-related data from production databases; user input.
Partners activities & responsibilities	
Partner owner of the device	The device will be owned to the industry, where the data collection is going to be performed.
Partner in charge of the data collection (if different)	SENSAP for the sensors. HOLONIX with i-Like machines at a higher level.
Partner in charge of the data analysis (if different)	CORE.
Partner in charge of the data storage (if different)	HOLONIX.
WPs and tasks	WP6/T6.1.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	The outputs of the machine learning systems will be accompanied by the timestamps along with naming of each individual value.
Standards, Format, Estimated volume of data.	The data will be stored in a local database and will also be provided via a REST api through the cloud. Estimated volume TBD according to aggregation processes.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The end-users will use the outputs to identify anomalous machine behaviour and expect future machine states.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Data sharing possible through HTTP RESTful web services.
Embargo periods (if any)	None
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data is going to persisted to relational database system. A regular back up service will run in the background, and aging algorithm will decide which records are to old and need to be removed.

10.5 SHADOW

Data Identification	
Data set description	Ros Bags – a bag is a file format in ROS for storing ROS message data.
Source (e.g. which device?)	Gripper.
Partners activities & responsibilities	
Partner owner of the device	No device still identified.
Partner in charge of the data collection (if different)	TBD.
Partner in charge of the data analysis (if different)	TBD.
Partner in charge of the data storage (if different)	Probably not need.
WPs and tasks	WP4/T4.4
Standards	
Info about metadata (Production and storage dates, places) and documentation?	N/A.
Standards, Format, Estimated volume of data.	N/A.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	N/A.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	GitHub repository will be available to partners of consortium.
Data sharing, re-use and distribution (How?)	N/A.
Embargo periods (if any)	None
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	TBD

10.6 HOLONIX

Data Identification	
Data set description	Data that will be stored in the KBS include real time data acquired by production machines/systems or additional sensors, including the electronic nose developed by SACMI, production related data from MES/ERP and additional data provided through UI.
Source (e.g. which device?)	Machines, production systems, MES/ERP, UI.
Partners activities & responsibilities	
Partner owner of the device	The device will be owned to the industry, where the data collection is going to be performed.
Partner in charge of the data collection (if different)	SENSAP (maybe also other various partners related to the specific incident and/or operation).
Partner in charge of the data analysis (if different)	BRUNEL, CORE, ATLANTIS.
Partner in charge of the data storage (if different)	• ATLANTIS will persist the data into a relational database. • HOLONIX will utilize that data to build the RDF data.
WPs and tasks	WP4/T4.5 – WP5/T5.3
Standards	
Info about metadata (Production and storage dates, places) and documentation?	Detailed Information about how to interact with HTTP Restful Web Services developed upon KBS will be provided.
Standards, Format, Estimated volume of data.	Data will be store in the KBS, made available to the other partner for further analysis and visualised in a UI.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	Data will be store in the KBS, made available to the other partner for further analysis and visualised in a UI.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Data sharing will be achieved through HTTP RESTful web services and/or message broker such as RabbitMQ.
Embargo periods (if any)	None
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data will be stored in servers owned by HOLONIX for the duration required by iQonic project

10.7 SENSAP

Data Identification	
Data set description	Production (shopfloor) data will be collected through ITK-Integra by utilizing its sensors. Moreover, ITK supports MQTT, Modbus, ProfiNet and OPC-UA (client) communication protocols for acquiring data from third-party sources. In addition, the vision inspection module will provide quality related data of the produced product.
Source (e.g. which device?)	The ITK integrates, two optical diffusion type sensors, two rotary encoders, one set of temperature/humidity sensors. Optionally, motor vibration/temperature sensor (up to four), energy consumption module. It integrates a vision inspection module.
Partners activities & responsibilities	
Partner owner of the device	SENSAP
Partner in charge of the data collection (if different)	Various partners related to the specific incident and/or operation.
Partner in charge of the data analysis (if different)	BRUNEL, CORE, ATLANTIS, HOLONIX
Partner in charge of the data storage (if different)	• ATLANTIS will persist the data into a relational database. • HOLONIX will utilize that data to build the RDF data.
WPs and tasks	WP3/T3.4 and WP4/T4.1.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	• Each produced data will be attached with a timestamp. • In general, the common model data will provide the metadata of the data source.
Standards, Format, Estimated volume of data.	The format of the data will follow the common data model (WP2, WP3).
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The ITK will collect, transform (to the canonical data model) and push the data to a message bus in order to be available to the upper layer.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The data is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Data sharing is capable through MQTT Pub/Sub pattern. Optionally, through HTTP RESTful web services.
Embargo periods (if any)	None
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	TBD.

10.8 HiLASE

Data Identification	
Data set description	LIDT measurement data having info on damage threshold values of crystals made by FILAR. Basic data on characterization of crystals as a part of simple laser systems (design of the systems not included). Data files with performance of thin-disk bonded during the project (no info on laser design).
Source (e.g. which device?)	Different devices for characterization of LIDT and laser performance clarification (beam profile, average power, pulse duration, energy, etc).
Partners activities & responsibilities	
Partner owner of the device	HiLASE
Partner in charge of the data collection (if different)	HiLASE
Partner in charge of the data analysis (if different)	FILAR and FRAUNHOFER.
Partner in charge of the data storage (if different)	• HiLASE will store info in the form of reports. • FILAR will collect files with info related to production procedures.
WPs and tasks	WP5/T5.5
Standards	
Info about metadata (Production and storage dates, places) and documentation?	The dataset might be accompanied with a documentation. Possible metadata include: • LIDT value • Performance of crystals as a part of simple laser systems
Standards, Format, Estimated volume of data.	The data will be stored at pdf format and are estimated to be at least 300MB per report.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The data will be used to decide whether or not a defective part/product should return to a previous production stage.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Cloud of the project.
Embargo periods (if any)	Depending on created data.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data is going to persisted to relational database system. A regular back up service will run in the background to ensure data reliability and to avoid data losses. Optionally, an aging algorithm may be used to decide which records are too old and need to be removed.

10.9 PRIMA

Data Identification	
Data set description	Real time data defined at various step in production line. Most appropriate are: • Vision images generated at Vision Inspection steps • Chip codes generated at the top side vision inspection of chip and chip on carrier • Optical beam generated at optical lenses alignment step • Lenses position generated at optical lenses alignment step
Source (e.g. which device?)	• Cameras; • Positions encoders
Partners activities & responsibilities	
Partner owner of the device	• PE department
Partner in charge of the data collection (if different)	• PE department
Partner in charge of the data analysis (if different)	• PE department • To be defined
Partner in charge of the data storage (if different)	• PE department
WPs and tasks	• WP8/T8.1 - PE as end user will demonstrate the IONIQ solution in its laser diode production systems
Standards	
Info about metadata (Production and storage dates, places) and documentation?	• Data are stored in SQL database. • All data (timestamp, numerical values, images, etc.) are stored in tables and linked logically
Standards, Format, Estimated volume of data.	• About 50Mb /day
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	• Data will follow product defining its position in product line and whether he can step forward (e.g. validation data, testing data, visual data, ..) as well as defining pass/fail criteria
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	• Data set is strictly confidential. Selected sub-set of data will be available only to the members of the consortium
Data sharing, re-use and distribution (How?)	• Selected sub-set of data will be mirrored on server, and accessible by FTP
Embargo periods (if any)	• none
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	• Data are back up at PE production line level

10.10 ALPES

Data Identification	
Data set description	• Wafer-mapping data from various inspection stages. • Wafer validation data including accelerated-lifetime tests. • Device performance data at various production stages. • Wafer fabrication process data.
Source (e.g. which device?)	Various.
Partners activities & responsibilities	
Partner owner of the device	ALPES.
Partner in charge of the data collection (if different)	ALPES.
Partner in charge of the data analysis (if different)	TBD.
Partner in charge of the data storage (if different)	TBD.
WPs and tasks	Various WP's and tasks that make use of data produced at end-users.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	All datasets will be accompanied by timestamp and associated metadata for the device, including production batch, wafer location and TBD relevant information. Real devices will be anonymized if necessary.
Standards, Format, Estimated volume of data.	ALPES uses open formats for storing / transferring the data. Estimated volume TBD.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	Finding correlations between data acquired by sensors and performance (or failure) of devices. Determine necessity of re-working; categorization of devices.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	Confidential, only for members of the Consortium and Commission services.
Data sharing, re-use and distribution (How?)	Managed through iQonic software solutions.
Embargo periods (if any)	None.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data will be streamed to the various iQonic software solutions which will take care of storage and backup. However, local copies of the raw data will be stored at ALPES as well and kept as long as the data have any use.

10.11 FORTH

Data Identification	
Data set description	Images (.dat, tiff)
Source (e.g. which device?)	Microscope
Partners activities & responsibilities	
Partner owner of the device	No device still identified.
Partner in charge of the data collection (if different)	No device still identified.
Partner in charge of the data analysis (if different)	Not identified yet.
Partner in charge of the data storage (if different)	Not identified yet.
WPs and tasks	WP4/T4.3
Standards	
Info about metadata (Production and storage dates, places) and documentation?	N/A.
Standards, Format, Estimated volume of data.	N/A.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	N/A.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	Data will be available to partners of consortium.
Data sharing, re-use and distribution (How?)	N/A.
Embargo periods (if any)	None.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Not identified yet.

10.12 SACMI

Data Identification	
Data set description	Real time data acquired by electronic olfactory system. These data include raw data from the chemical sensors, environmental parameters (temperature, humidity), output elaborated by the instrument (classification and quantification of the odour detected). Data will be used to detect the presence of contaminants and decide future actions.
Source (e.g. which device?)	Electronic olfactory system.
Partners activities & responsibilities	
Partner owner of the device	SACMI
Partner in charge of the data collection (if different)	SACMI, SENSAP, use-case where the device is installed
Partner in charge of the data analysis (if different)	SACMI, SENSAP, BRUNEL, CORE, HOLONIX, ATLANTIS.
Partner in charge of the data storage (if different)	TBD
WPs and tasks	WP4/T4.2
Standards	
Info about metadata (Production and storage dates, places) and documentation?	The dataset might be accompanied with a documentation. Possible metadata include: • Date and time. • Calibration parameters of the chemical sensors. • Environmental and diagnostic parameters (temperature, humidity, gas flow, etc.).
Standards, Format, Estimated volume of data.	The data will be stored as text files or excel sheets and are estimated to be no more than 10MB per day.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The data given by the electronic olfactory system will be used to detect the presence of volatile organic compounds related to the possible presence of defects in the optoelectronic assembly. These data will be collected together with SENSAP sensing platform. The RSC will then use the data to decide whether or not a defective part/product should return to a previous production stage.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	TBD.
Embargo periods (if any)	None.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data can be stored inside the hard drive of the electronic olfactory system and/or shared with an external database.

10.13 FICONTEC

Data Identification	
Data set description	During the machine assembly and process development some first date sets will be created at FICONTEC in house. For these datasets the fundamental structure and way how to save and process the data from the assembly machine towards the project partners is the main goal. Typical sets of data comprise information about process parameters and related processing results. In addition, many of the data generated will originate from the newly developed sniffing, inspection and handling tools from partners. Hence, no generalized description can be given.
Source (e.g. which device?)	Data sources are devices that are used during the whole process chain: • Inspection data obtained during image recognition. • Alignment (data of active and passive alignment measurement devices). • Attachment (pre- and post-bond alignment state, bonding strength). • Optical characterization. • Mechanical sensing (e.g. by a SHADOW robot handling tool). • Chemical information (e.g. via the sniffing tool).
Partners activities & responsibilities	
Partner owner of the device	FICONTEC.
Partner in charge of the data collection (if different)	FICONTEC, PRIMA with help of partners owing the individual hardware.
Partner in charge of the data analysis (if different)	BRUNEL, CORE, ATLANTIS, HOLONIX.
Partner in charge of the data storage (if different)	TBD with technical partners.
WPs and tasks	WP3/T3.4 and WP4/T4.1
Standards	
Info about metadata (Production and storage dates, places) and documentation?	• Each produced data can be attached with a timestamp. • Each produced data can be attached with a parameter stamp. • TBD with technical partners: unified parameter.
Standards, Format, Estimated volume of data.	The format of the data can follow the common data model (WP2, WP3). An OPC UA server is under discussion.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	One the one side FICONTEC is providing the hardware (machine interface) to read out data from individual hardware installed in the assembly machine. On the other hand, FICONTEC is generating own data via the sensor systems build into the machine.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The data is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Data sharing within the consortium is allowed. Data sharing and re-distribution outside the consortium requires confirmation by FICONTEC, as data might comprise proprietary information.
Embargo periods (if any)	TBC
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	TBC.

10.14 TAU

Data Identification	
Data set description	Various sensors and quality control systems can generate non-real time data. A relatively large amount of data can be generated before and after the assembly stage, stored, and analysed to improve the reliability and quality control of the (sub)assemblies. The data may be generated by different equipment in the process chain.
Source (e.g. which device?)	The data is generated by sensors and imaging systems.
Partners activities & responsibilities	
Partner owner of the device	The partner generating the data.
Partner in charge of the data collection (if different)	SENSAP.
Partner in charge of the data analysis (if different)	SENSAP.
Partner in charge of the data storage (if different)	TBD.
WPs and tasks	WP5-WP6.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	The dataset can be stored with a content documentation. The metadata may include information about the production chain and parameters, timestamps, etc.
Standards, Format, Estimated volume of data.	Not defined yet.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The will be used to optimize the process chain and improve the quality control of micro-optic (sub)assemblies. The collected data enable to detect defects before and after the assembly stage. The generated non-real time data may be essential to identify long-term reliability issues, reduce rework costs, and improve the performance of (sub) assemblies.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The full dataset will be confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Most suitable protocols for data sharing are RabbitMQ and HTTP Restful webservices but data sharing is yet to be decided along with the industrial partners.
Embargo periods (if any)	None
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	TBC.

Data Identification	
Data set description	Simulation data will be generated to support the assembly and process chains.
Source (e.g. which device?)	The data is generated by simulation software and hardware.
Partners activities & responsibilities	
Partner owner of the device	TAU but not necessary, maybe also other various partners that generate simulation data. (e.g. BRUNEL).
Partner in charge of the data collection (if different)	TBD.
Partner in charge of the data analysis (if different)	F-IOF.
Partner in charge of the data storage (if different)	The partner generating the data.
WPs and tasks	WP3.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	The data can be accompanied by the documentation. The metadata will be defined later.
Standards, Format, Estimated volume of data.	The used data formats are not defined yet.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The data will be used to analyse and optimize the assembly process chain.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential and available only to the members of the consortium.
Data sharing, re-use and distribution (How?)	Data sharing is capable through HTTP RESTful web services and RabbitMQ Pub/Sub pattern.
Embargo periods (if any)	None.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	The data storage is not decided yet.

10.15 BRIGHTERWAVE

Data Identification	
Data set description	Characterization data in the incoming quality inspection before the assembly as well as during the fabrication process of the key components used in the assembly. The DSS system can be used to define the severity of the defects and decide future actions both during the component fabrication and in the incoming quality inspection. The real time data (such as optical output power) during the assembly process. Data is used to optimize the positions of the components used in the assembly.
Source (e.g. which device?)	Optical spectrum analyser and L-I-V (light-current-voltage) measurement system.
Partners activities & responsibilities	
Partner owner of the device	BRIGHTERWAVE's industry supply chain partners and assembly partner in Asia/Europe where the data collection is going to be performed.
Partner in charge of the data collection (if different)	BRIGHTERWAVE.
Partner in charge of the data analysis (if different)	BRIGHTERWAVE.
Partner in charge of the data storage (if different)	BRIGHTERWAVE.
WPs and tasks	WP8/T8.4.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	The dataset might be accompanied with a documentation. Possible metadata include at least date, time, component ID, measurement system and measurement parameters (i.e. wavelength range used in the measurement).
Standards, Format, Estimated volume of data.	The data will be in ASCII format and is estimated to be at least 10MB per day.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	The DSS in the incoming quality inspection as well as in the component fabrication will use the data to decide whether or not a defective part should be rejected or is it possible to re-use the part for a different product (i.e. product with different characteristics / operation conditions). In addition, the data can be used the optimize the yield of the fabrication process and the assembly process in the following fabrication rounds. Furthermore, the data can be used to optimize the component design in the following device iteration rounds.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential. The part of the data is available to the members of the consortium, but the part of it is not shared even within the consortium.
Data sharing, re-use and distribution (How?)	TBD.
Embargo periods (if any)	TBC.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	The data will be stored at BRIGHTERWAVE's server. Backup of the data will be taken regularly. The data won't be removed.

10.16 CRIT

Data Identification	
Data set description	CRIT won't create an industrial data set. Our dataset will collect general information about partners and project's activities, as well as information about news, events, publications relevant for the project. Moreover, we will collect statistical data on users' and general public access to the website and social networks.
Source (e.g. which device?)	Direct contact from partners/Software for statistical analysis (Excel).
Partners activities & responsibilities	
Partner owner of the device	N/A.
Partner in charge of the data collection (if different)	N/A.
Partner in charge of the data analysis (if different)	N/A.
Partner in charge of the data storage (if different)	N/A.
WPs and tasks	WP9/ T9.2.
Standards	
Info about metadata (Production and storage dates, places) and documentation?	TBD.
Standards, Format, Estimated volume of data.	Data will be used for the communication and dissemination of the project activities and of partners role in the project. Moreover, we will monitor users' access to the website and social network pages to evaluate stakeholders' interest in the project activities.
Data exploitation & sharing	
Data exploitation (purpose/use of the data analysis)	Data regarding partners, project activities, news, events are open; data about users' access will be shared with project partners. Data on publications will be made open only after non-disclosure breaking checks.
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	The dataset is confidential. The part of the data is available to the members of the consortium, but the part of it is not shared even within the consortium.
Data sharing, re-use and distribution (How?)	TBD.
Embargo periods (if any)	Embargo periods (more or less 25 days) will be guaranteed for double partners' check on the content of the publication.
Archiving & preservation (including storage and backup)	
Data storage (including backup): where? For how long?	Data will be stored in internal repositories for at least 5 years after the project's end.