DIGITAL MANUFACTURING PLATFORMS FOR CONNECTED SMART FACTORIES

D1.7 Data Management Plan

Deliverable Id:	**D1.7**
Deliverable Name:	**Data Management Plan**
Status:	**Delivered**
Dissemination Level:	**PU**
Due date of deliverable:	**30/06/2019**
Actual submission date:	**02/07/2019**
Work Package:	**WP1**
Organization name of lead contractor for this deliverable:	**INNOVALIA**
Author(s):	**INNO**
Partner(s) contributing:	**UNPARALLEL, ATB, ATLANTIS, CEA tech, CONTINENTAL, DANOBAT, JSI, MONDRAGON, PHILIPS, PRIMA, SINTEF, TU BS, TELEFONICA I+D, TTTECH, THYSSENKRUPP, UNIMETRIK, VTT, WHIRLPOOL.**

Abstract : The Data Management Plan is an Open Research Data Pilot (ORDP). This deliverable is an outcome from of T1.3. Pilots Coordination and it was illustrated in Section 2.

This deliverable has been developed for ensuring an appropriate quality management of data collection and sharing due to it is essential for QU4LITY Project

www.QU4LITY-project.eu

Co-funded by the Horizon 2020 Programme of the European Union

Contents

HISTORY

Version	Date	Modification reason	Modified by
0.1	12/05/2019	Table of contents and first draft	Carmen Polcaro (INNO)
0.2	31/05/2019	Definition of the requirements of the datasets	Tiago Teixeira, Bruno Almeida (UNPARALLEL)
0.3	06/06/2019	Collection of datasets	Tiago Teixeira, Bruno Almeida (UNPARALLEL)
0.4	25/06/2019	Integration of the datasets	Carmen Polcaro (INNO)
0.5	28/06/2019	Executive summary and conclusions	Carmen Polcaro (INNO)
0.6	01/07/2019	Final version for reviewing	Carmen Polcaro (INNO)
0.7	01/07/2019	Review and comments	Diego Esteban (ATOS)
1.0	02/07/2019	Final version for submitting	Carmen Polcaro (INNO)

1. Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 825030, and for this reason we are gratefully acknowledge. Also, we would like to thank the QU4LITY consortium for providing feedback on the drafts of this deliverable. The content of this documents reflects only the authors' view and INNOVALIA is not responsible for any use that may be made of the information it contains.

2. Executive Summary

The present report is focused on the preparation of the Data Management Plan (DMP) for QU4LITY project. DMP provides an analysis of the main elements of the data management policy that will be used throughout the project with regard to all the datasets that will be generated. In particular, DMP defines what data is collected and/or generated, how this data will be managed and shared by the project partners, and also, how this information will be updated and preserved during and after the project duration.

DMP of QU4LITY project describes the life cycle of all modelling and observation data collected and processed in the project, giving an overview of available research data, access, the data management and terms of use. The DMP reflects the current state of the discussions, plans and ambitions of the partners, and will be updated and augmented with new datasets and results during the lifespan of QU4LITY.

The Open Research Data Pilot of the European Commission (EC) aims to improve and maximize access to and re-use of research data generated by projects focusing on encouraging good data management as an essential element of research best practice. Following the recommendation of the EC, QU4LITY project is participating on the Open Research Data Pilot and DMP is included as a deliverable (D1.7, Month 6) in charge of WP1.

3. Introduction and Background

Research and innovation projects such as QU4LITY used to produce large amounts of data. Depending on the discipline, this data could come for several ways, for example from social science research, laboratory testing, field studies or observations. However, it often remains unclear and uncertain what will happen with the data after they were analysed and the project was finished.

Moreover, a lot of data sets are interesting for other researchers, although due to the fact that they are either stored on a local server or miss crucial-metadata, their potential value cannot be exploited.

Hence, researcher need to think about the data they will produce at the beginning of the research. Due to this reason the Data Management Plan (DMP) must be created.

The purpose of this deliverable is to provide an analysis of the main elements of the data management policy that will be used during the development of QU4LITY project as well as by the whole consortium regarding to the project research data. This document covers the complete research of data life cycle and describes the different types of data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for its verification or reuse.

It is expected that the DMP will be a living document, which will evolve during the lifespan of the project, particularly whenever significant changes arise such as dataset updates or changes in Consortium policies.

This is the 1st version of the DMP and it is delivered in M6. The document includes an overview of the datasets produced by the project and the specific conditions that are attached to them.

Although this 1st version contains a broad range of aspects related to the management of data during the project, in later versions it will be get into more detail in particular issues such as data interoperability and practical data management procedures, mainly those collected in the pilots developed in QU4LITY.

This documents account with several sections as is required for all H2020 projects and is suggested in the template of the Data Management Plan provided by European Commission.

The 1st section, Data Summary, provides and overview of the data set that will be produced during the QU4LITY project. Moreover, it describes the origin of the data as well as the format and the allocation of particular WPs, the purpose of the collection as well as information on the utility.

The 2nd section, FAIR data, clearly points out, which data will be made openly accessible and which won't. The choice of one or another option will be clearly justified. This is quite relevant for the primary data that will be collected as part of the first tasks developed in QU4LITY. Other information in this section is data repositories or other locations where the data will be stored.

The 3rd section, Allocation of Resources, highlights aspects related to the cost of the accessibility.

In 4th section, will be discussed aspects such as security of the data generated during the QU4LITY project.

In the 5th section the main ethical issues will be presented.

As for the two final sections, both are very general and are for issues that have not been treated before as well as to give support to the development and updating of the Data Management Plan.

During the preparation of this document a very important issue as DATA MANAGEMENT POLICY should be considered because since collecting and processing Zero Defect Manufacturing datasets (for example quality or management data) is essential for the project and it is of paramount importance to ensure an appropriate quality management of data collection and sharing. As part of the establishment and operation of the project's use cases QU4LITY will produce and make available datasets in line with European privacy and data protection regulations. Selected datasets will be made available as open data in order to enable its wider use (for example as part of the QU4LITY market platform)

Those deliverables of QU4LITY project that are related to the use cases coordination will include the project's data management plan (DMP) concerning data sets at these sites. Moreover, a data manager will be identified for each pilot and its main responsibilities will be the provision with relevant guidelines and documentation for data management.

QU4LITY will produce and make available datasets, some in line with privacy and data protection validations, concerning production operations as part of the testbed validation of project's results. QU4LITY data will be a critical asset for the manufacturers of the consortium in general and for the pilot participants in particular. A plan for managing these assets in line with the IPR strategy of the project will be devised as part of the "Exploitation and Sustainability Plan" in WP9.

Finally, the open access to the scientific publication will be made in a way that the open access to the peer reviewed scientific publications of the project will be provided with the highest standard (gold) when possible. The WP9 will have the key role of keeping track of the consortium publications and ensuring their dissemination.
That QU4LITY project accounts with a Data Management Plan is essential at least for an efficient and properly management of the pilot's data, optimise the present interaction as well as the adaptation of the data collected to the future organizational needed, and of course for accomplishing with legal requirements.

The first task is the identification of the requirements for the DMP, and it should interact in the following areas that make possible to reach a unique vision as well as guarantee the integrity, completeness accuracy and reliability of the whole data generated in QU4LITY because in this way it will be act in all the life cycle and the font of data available for sharing its information:

Data Management Area Number	Data Management Area Name
1	Data Integration
2	Data Quality
3	Data Governance
4	Accessibility to the information
5	Management of Key/Master Data
6	Metadata management
7	Data Mining
8	Performance Management
9	Data security

Other aspects that should be considered in the DMP preparation are the companies that will be benefited from DMP, the advantages of applying it, the things that are necessaries for the realisation of the DMP, the priorities of it, the person in charge of the design and implementation, the tools available and needed, the way for ensuring the quality and security of the data, the way for updating the plan and which KPIs will allow tracking.

Additional to aspects that should be considered there are several questions that will be solved thanks to the application of the DMP such as data discovering, capture and migration; development and management of databases; data maintenance; interaction with databases (DB query), statistics analysis, relational databases and data analysis.

For the development of the Data Management Plan two phases must be considered the design and the implementation of the plan.

Regarding to the Data Management Plan Design there are a series of components mandatories like the purpose description and scope of the Plan, the data description (format and nature), the rules for applying within the format (metadata,…), the storage plans, legal aspects (For example IPR or confidentiality), policies of data access and restrictions establishment, assignation of the responsibilities on the data management (roles, competences, guaranties and supervision), big data treatment and finally technological platform section (the best technological platform should be selected). In addition to the previous, a sustainable strategy should be design for picking up the key points and that help the partners involved in the creation of the DMP as the started point. In particular, scalability must be guarantee as well as ensure a real time cleaning, validation, integration and data enrichment efforts for its availability in the future.

In case of Development and Implementation of Data Management Plan several steps should be followed for the development among which are in case of strategy the vision and mission, goals and objectives, alignment with the organisational strategy and also with the business strategy and finally the guiding principles; in case of the organisation the operational model must be considered as well as the participants with its role and responsibilities and the data property; in case of policies, standards

and process it must be considered the rules and policies, process and controls, the data definition, the data standards and metadata classification and taxonomy; for the monitoring and evaluation the Statistics and analysis, Progress monitoring, Problems detection, Continuous improvement, Score-carding; for technology it has to be considered the architecture, quality and security of data and the metadata repository and finally in case of communication the communication plan, the individual updating and the training.

When implementing the DMP it is recommended a fluent communication as well as an adequate information addressed to all business users. It is also essential to provide a coordination mean between partners because this is the only way for conflicts reduction and for boosting the consensus. The most critical step is the data requirement due to it guarantee the viability of DMP. Finally, it is necessary to move forward with slow and secure steps. As a protection measure, the development of a transition plan must be considered so it can work until the maximum efficient point needed will be reach.

Finally, in the planification, it must be considered the change management for measuring the DMP impact in the business, the preparation of the communication strategy and account with the support of IT department.

3.1 Participation in the pilot on open research data

The European Commission (EC) is running a flexible pilot under Horizon 2020 call the Open Research Data Pilot (ORD pilot) whose aim is improving and maximising access to and re-use of research data generated by H2020 projects and takes into account the need to balance openness and protection of scientific information, commercialisation and intellectual property rights (IPR), privacy concerns, security as well as data management and preservation questions. The 2017 work programme of ORD pilot has been extended to cover all the thematic areas of Horizon 2020.

Following the recommendation of the EC, QU4LITY project is participating in the Open Research Data Pilot and the DMP is considered as a deliverable (D1.7) due in month 6. DMP of this project has been prepared by taking into account the document template of the "Guidelines on Data Management in Horizon 2020" (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf). This document will be updated and augmented with new datasets and results according to the progress of activities of QU4LITY project. Also, DMP will be updated including changes in consortium composition and policies over the course of the project.

The procedures that will be implemented for data collection, storage and access, sharing policies, protection, retention and destruction will be according to the requirements of the national legislation of each partner and in line with the EU standards.

3.2 Building a DMP in the context of H2020

The EC provides a document with guidelines for project participants in the pilot. The guidelines address aspects like research data quality, sharing and security. According to the guidelines, projects participating will need to develop a DMP. This document has been produced following these guidelines and aims at providing a consolidated plan for QU4LITY partners in the data management plan policy that the project will follow.

The consortium will comply with the requirements of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. The consortium will preserve the right to privacy and confidentiality of data of the participants in the survey by providing all participants to the survey with two documents: The Participant Information Sheet and the Consent Form. These documents will be sent electronically and will provide information about how the answers will be used and what is the purpose of survey.

The participants will be assured that their answers will be used only for the purposes of the specific survey. The voluntary character of participation will be stated explicitly in the Consent Form. The consortium will examine before conducting the survey following the requirements of the national legislation in line with the EU standards whether the proposed data collection requires special local/national ethical/legal permission.

An ethical approach will be adopted and maintained throughout the fieldwork process. The responsible partners will assure that the EU standards regarding ethics and Data Management are fulfilled. Each partner will proceed with the survey according to the provisions of the national legislation that are adjusted according to the respective EU Directives for Data Management and ethics.

The recruitment process to be followed by the consortium for the engagement of stakeholders (including inclusion/exclusion criteria for all the surveys) will be transparent and such criteria will be included and explained in the Participant Information Sheet.

Participants to the survey will be invited by each partner by email. The third parties that will be invited to participate in the survey will have no role in QU4LITY and no professional relationship with the consortium. The consortium will also examine whether personal data will be collected and how to secure the confidentiality in such a case.

The Steering Committee of the project will also ensure that EU standards are followed. The issue of informed consent for all survey procedures, all participants will be provided with a Participant Information Sheet and Consent Form to provide informed consent. The default position for all data relating to residents and staff will be anonymous.

4. DATA MANAGEMENT PLAN (DMP)

4.1 General Description

This document consists of the first version of the project's DMP, that is contained as a deliverable (D1.7, M6) in charge of WP1 Project Management and Administration.

The main purpose of a DMP is to provide and analysis of the main elements of the data management policy that will be used by the consortium with the regard to all the datasets that will be generated during the project development.

This document describes the Research Data with the metadata attached and overview of the datasets to be produced by the project, their characteristics and their management processes to make them discoverable, accessible, assessable, usable beyond the original purpose and exchangeable between researchers. It also presents the specifications of the dedicated Data Management Portal developed by the project in the context of the Open Research Data Pilot, allowing the efficient management of the project's datasets and providing proper open access on them for further analysis and re-use. In addition, DMP of QU4LITY project reflects the current status of discussion within the consortium about the data that will be produced.

4.2 Activities of Data Management Plan

The DMP is a dynamic document, updated throughout the whole project lifecycle. The final version of this report will be delivered before the end of the project (M36), reflecting on lessons learnt and describing the plans implemented for sustainable storage and accessibility of the data, even beyond the project's duration.

A Knowledge Management system will be developed, which incorporates, in a structured way, the technical and business knowledge created during the project. The activities of the QU4LITY concerning the data management are planned as follows:

- Knowledge management – to be led by the Dissemination and Exploitation Manager (DEM), in which the DMP will be delivered.
- A knowledge management document will be created, based on DMP, describing how the acquired data and knowledge will be shared and/or made open, and how it will be maintained and preserved. The identifiable project data will be provided in a manner to define the relevant knowledge, increase partners' awareness, validate the result, and timeframe of actions.
- Technology watch - 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.

4.3 Register of datasets generated or collected in QU4LITY

The intention of the DMP is to describe the different kind of datasets collected or created by QU4LITY activities during the runtime of the project. The register of datasets has to be understood as living document, which will be updated regularly during project lifetime. The project started in January 2019, so currently (M6) there is no dataset generated or collected.

The information listed below reflects the conception and design of the individual partners in the different work packages at the beginning and mid-term of the project. The data register will deliver information according to information detailed in Annex 1 (Part A) of the Grant Agreement Document (GA):

- Data set reference and name: identifier for the data set to be produced.
- Data set description: descriptions of the data that will be generated or collected, its origin or source (in case it is collected), nature, scale, to whom it could be useful and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
- Partners activities and responsibilities: partner owner of the device, in charge of the data collection, data analysis and/or data storage, and WPs and tasks it is involved.
- Standards and metadata: reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created. Format and estimated volume of data.
- Data exploitation and sharing: description of how data will be shared, including access procedures and policy, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.) and if this information will be confidential (only for members of the Consortium and the Commission Services) or public. In case of dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).
- Archiving and preservation (including storage and backup): description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

Each of these sections cover the main principles of FAIR data, that is, the research data should be "FAIR": findable, accessible, interoperable and re-usable. In this context, QU4LITY partners have been identified their implementation choices regarding their research data.

4.4 Metadata for Data Management

Partners have characterised their research data and associated software and/or used in the project whether these are:

1. Findable/Discoverable; Can be discovered by means of an identification mechanism such as Digital Object Identifier.
2. Accessible: Define in what modalities, the scope of the action, stablish the licenses and define the IPR
3. Assessable and Intelligible: Allow to third parties to make assessments.
4. Useable beyond the original purpose for which it was collected or usable to third parties after the collection of the data for long periods (repositories, preservation and curation)
5. 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.

4.5 Data description

In order to collect the information about the research data that will be generated in QU4LITY from the different activities of the project, we have elaborated a template to be completed by the consortium partners. This template includes the following information items:

- Dataset reference and Name: name, homepage, publisher and maintainer.
- Dataset description: description, provenance, usefulness, similar data, re-use and integration
- Standards and metadata: metadata description, vocabularies and ontologies.
- Data sharing: License; URL dataset description, openness, software necessary, repository
- Archiving and preservation: preservation, growth, archive and size

- ### 4.5.1 Dataset per partner

The following partners have identified the data that will be produced in different project activities, these datasets will be updated and other datasets will be added during the project.

Table 1 – Dataset per partner

CEA tech	
Data Identification	
Dataset name and description	Acoustic sensor data
Pilot	ThyssenKrupp, Kolector
Description and metadata	
Datatype	Experimental Data - Data derived from laboratory equipment that is subject to controlled conditions
Format	.TXT
Size	100MB
Info about metadata (production and storage data, places) and standards	Test facility, test number, date, time, sensor number.
Data access and sharing	
Access to data	Public
Reason for restricted access	
Requirements to access	
License for datasets	
Data sharing	Allowed
Reason for unavailable sharing	
Archiving and preservation	
Data storage (including backup)	
Period of archiving	

TECHNISCHE UNIVERSITÄT BRAUNSCHWEIG	
Data Identification	
Dataset name and description	Siemens: Production process data for zero defect manufacturing

Pilot	Siemens, Continental
Description and metadata	
Datatype	Observed Data - Unique data that is collected in real time and cannot be reproduced
Format	.XML, .TXT, .CSV
Size	>2GB
Info about metadata (production and storage data, places) and standards	Prediction results as numbers
Data access and sharing	
Access to data	Restricted
Reason for restricted access	Permission to access must be granted by the pilot owner
Requirements to access	
License for datasets	
Data sharing	Not allowed
Reason for unavailable sharing	Permission for sharing must be granted by the pilot owner
Archiving and preservation	
Data storage (including backup)	NA
Period of archiving	NA

VTT	
Data Identification	
Dataset name and description	User needs for Equipment Solutions Marketplace
Pilot	Not related to pilots: The material consists of interview on the needs and requirements of industrial and other partners. The information will be used to prepare the concept and specifications for the Equipment Solutions Marketplace.
Description and metadata	
Datatype	Data collected through interviews
Format	.DOC
Size	300KB
Info about metadata (production and storage	NA

data, places) and standards	
Data access and sharing	
Access to data	Restricted
Reason for restricted access	The interview documents are confidential and will be used only by the persons involved in Task 8.3.
Requirements to access	Project internal document storage and management tool
License for datasets	
Data sharing	Not allowed
Reason for unavailable sharing	The interview documents are confidential and will be used only by the persons involved in Task 8.3.
Archiving and preservation	
Data storage (including backup)	Project internal database
Period of archiving	During the project duration

CONTINENTAL	
Data Identification	
Dataset name and description	Lot Release Risk Evaluation Data Evaluate the risk for failure sitting in a batch before shipment. All parts ready for shipment are tested ok. Still there might be parts from nearby batches, that show defects or trends that could lead to defects. Data collection happens via test and inspection equipment connected to the Continental MES system. In addition as well environmental condition sensors might be used. Storage happens conventionell in Database and as well in the cloud.
Pilot	Continental
Description and metadata	
Datatype	Derived Data - Data that is a result of the analysis of data or aggregates from various sources
Format	unstructured data, .CSV
Size	>1TB
Info about metadata (production and storage data, places) and standards	Trend data, correlations, Results from application of AI
Data access and sharing	

Access to data	Restricted
Reason for restricted access	Data distribution outside the consortium might cause issues regarding competition laws.
Requirements to access	Business Partner access might be given and is common Procedure in Continental
License for datasets	No license
Data sharing	Allowed
Reason for unavailable sharing	
Archiving and preservation	
Data storage (including backup)	Continental Cloud, Continental MES system
Period of archiving	15 Years

JOSEF STEFAN INSTITUTE	
Data Identification	
Dataset name and description	Kolektor Pilot dataset Robotic motion data for quality inspection, obtained during the implementation of robotic motion. the motion will be implemented to bring cameras at the correct location for inspection. It will be actual robot data and (most likely) a report.
Pilot	Kolector
Description and metadata	
Datatype	Experimental Data - Data derived from laboratory equipment that is subject to controlled conditions
Format	.PDF, .URDF, .BAG
Size	1 GB
Info about metadata (production and storage data, places) and standards	
Data access and sharing	
Access to data	Restricted
Reason for restricted access	It will be closely related to the Kolektor production line, and therefore confidential. it will also be very specific.
Requirements to access	It will be in common data formats for robotics, it will also have a short report (most likely)
License for datasets	

Data sharing	Not Allowed
Reason for unavailable sharing	The data will be specific to the implementation
Archiving and preservation	
Data storage (including backup)	
Period of archiving	

UNIMETRIK	
Data Identification	
Dataset name and description	Dimensional quality control data Dimensional quality control data gathered as a result of measurement of the aeronautical parts during their production on the machine tool with the implementation of the M3 software
Pilot	GF
Description and metadata	
Datatype	Observed Data - Unique data that is collected in real time and cannot be reproduced
Format	.PDF, .XML, .M3, .DMO, .DMI , .CSV, QIF
Size	aprox 1-300kb
Info about metadata (production and storage data, places) and standards	Standard: Quality Information Framework (QIF) system
Data access and sharing	
Access to data	Restricted
Reason for restricted access	Restricted access due to intellectual property of the end user
Requirements to access	It will be in common data formats for robotics, it will also have a short report (most likely)
License for datasets	M3 software
Data sharing	Allowed
Reason for unavailable sharing	
Archiving and preservation	
Data storage (including backup)	Not yet defined

Period of archiving	Not yet defined

THYSSEKRUPP	
Data Identification	
Dataset name and description	TKPF: ProductionData This dataset will provide production line values for our processes (screwing with a torque/angle curve and press-fitting with a force/distance curve) as well as acoustic control data (temporal raw accelerometers values).
Pilot	ThyssenKrupp
Description and metadata	
Datatype	Observed Data - Unique data that is collected in real time and cannot be reproduced
Format	.XML
Size	~12KB for each processes module file / ~3MB for each control module file
Info about metadata (production and storage data, places) and standards	
Data access and sharing	
Access to data	Restricted
Reason for restricted access	This represents our added value for TK assembly production, this data cannot be accessible by anyone and have to be restricted.
Requirements to access	XML, we will transform it to make it universal
License for datasets	
Data sharing	Not Allowed
Reason for unavailable sharing	ThyssenKrupp Quality policy does not allow the sharing.
Archiving and preservation	
Data storage (including backup)	Every weeks TK makes an image disk of our production line computers, everyday the central server are saved and an archive available for 15 years is made
Period of archiving	"Quality data: 1 year on Quality drive, 15 years on archive Process curves: 1 year on Quality drive Audio curves (Control) : 0,5 year on Quality drive, 1 year on archive"

ATLANTIS ENGINEERING	
Data Identification	
Dataset name and description	DSS ZDM Decision Support System dataset: receives processed data from heterogeneous sources. Produces recommendations, KPIs, suggestions for mitigation actions, triggers the activation of the appropriate ZDM strategy(ies).
Pilot	Mondragon1, Prima, Danobat
Description and metadata	
Datatype	Derived Data - Data that is a result of the analysis of data or aggregates from various sources
Format	.XML, .JSON
Size	~300MB per day
Info about metadata (production and storage data, places) and standards	"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"
Data access and sharing	
Access to data	Restricted
Reason for restricted access	Sensitive data and information from the pilot partners.
Requirements to access	Authentication and authorisation will be required. Relevant documentation will be provided.
License for datasets	
Data sharing	Allowed
Reason for unavailable sharing	
Archiving and preservation	
Data storage (including backup)	Data is going to archived and preserved in a relational database system.
Period of archiving	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.

ATLANTIS ENGINEERING	
Data Identification	
Dataset name and description	Prediction of defects, based on asset's deterioration Make use of data from the shopfloor (sensors, MES, CMMS and otrher systems) to assess the health status of an asset and its deterioration rate. A deteriorating asset is likely to operate in less than optimal conditions, thus diverging from the quality goals and the objective of ZDM.
Pilot	Mondragon1, Prima, Danobat
Description and metadata	
Datatype	Derived Data - Data that is a result of the analysis of data or aggregates from various sources
Format	.JSON
Size	~50MB per day
Info about metadata (production and storage data, places) and standards	Remaining Useful Lifetime (RUL), Degradation analysis, Run-to-failure analysis, Fault prediction
Data access and sharing	
Access to data	Restricted
Reason for restricted access	Sensitive data and information from the pilot partners.
Requirements to access	Authentication and authorisation will be required. Relevant documentation will be provided.
License for datasets	propietary
Data sharing	Allowed
Reason for unavailable sharing	
Archiving and preservation	
Data storage (including backup)	Locally at pilot sites
Period of archiving	Depends on the end user needs.

FRAUNHOFER ILT	
Data Identification	
Dataset name and description	Survey Results

	Results from survey on standards and norms
Pilot	No
Description and metadata	
Datatype	Derived Data - Data that is a result of the analysis of data or aggregates from various sources
Format	.xls
Size	1MB
Info about metadata (production and storage data, places) and standards	Meta data are created by the survey provider (LamaPoll)
Data access and sharing	
Access to data	Public (D2.7, D2.8)
Reason for restricted access	
Requirements to access	Any software that can handle .xls files will have access to these files
License for datasets	Propietary
Data sharing	Allowed
Reason for unavailable sharing	
Archiving and preservation	
Data storage (including backup)	After the end of the Project, raw data will be transferred into FhG Archives and are available for at least 10 years.
Period of archiving	10 yers

SINTEF	
Data Identification	
Dataset name and description	Philips pilot interview Action research on the line to Philips for connecting data on shopfloor.
Pilot	Philips
Description and metadata	
Datatype	Observed Data - Unique data that is collected in real time and cannot be reproduced

Format	Taped and transcripted
Size	
Info about metadata (production and storage data, places) and standards	
Data access and sharing	
Access to data	Restricted
Reason for restricted access	Internal IPR and GDPR
Requirements to access	NDS application https://nsd.no/
License for datasets	
Data sharing	Not allowed
Reason for unavailable sharing	Internal IPR for Philips
Archiving and preservation	
Data storage (including backup)	Transcripted into documents, internal Teamsite for Philps
Period of archiving	3-6 months

TELEFONICA I+D	
Data Identification	
Dataset name and description	Digital infrastructures test results Digital infrastructures test results will compile the data obtained during the testing phase of digital enablers platforms.
Pilot	Related to IT enablers tested in experimental facilities.
Description and metadata	
Datatype	Experimental Data - Data derived from laboratory equipment that is subject to controlled conditions
Format	.csv or others
Size	TBD
Info about metadata (production and storage data, places) and standards	TBD
Data access and sharing	

Access to data	Restricted
Reason for restricted access	Restricted to Qu4lity consortium
Requirements to access	
License for datasets	
Data sharing	Allowed
Reason for unavailable sharing	
Archiving and preservation	
Data storage (including backup)	TBD
Period of archiving	TBD

ATB	
Data Identification	
Dataset name and description	Context Repository The Context Repository will contain data from a production line/machine, which is used to identify the current context under which a production line/machine is operated. The identified context, which is derived from the observed data, can be used, by other software services, to identify for example anomalies, which could lead to maintenance activities.
Pilot	Continental
Description and metadata	
Datatype	Observed Data - Unique data that is collected in real time and cannot be reproduced
Format	RDF stored in a relational database
Size	TBD
Info about metadata (production and storage data, places) and standards	The data of the context repository contains observed raw data instantiated in an ontology, which will be specifically created for the Continental pilot within the QU4LITY project. The ontology will be documented.
Data access and sharing	
Access to data	Restricted
Reason for restricted access	Due to legal issues, the data will be shared among the Qu4lity consortium partners only. However, the structure of the Context Model (ontology) will be made available in public.

Requirements to access	Data can be accessed using Protege (open source Ontology editor) or via the Context Extractor (background of ATB).
License for datasets	
Data sharing	Not allowd
Reason for unavailable sharing	Due to legal issues, the data will be shared among the Qu4lity consortium partners only. However, the structure of the Context Model (ontology) will be made available in public.ed
Archiving and preservation	
Data storage (including backup)	Continental premises
Period of archiving	TBD

PHILIPS	
Data Identification	
Dataset name and description	Philips Dataset "Data contains: Machine data (status,errors) Process data (force, way, moulding, welding etc. etc.) Maintenance records Operators records Measurement values DMC codes Carrier data
Pilot	Philips
Description and metadata	
Datatype	Observed Data - Unique data that is collected in real time and cannot be reproduced
Format	Timeseries data, combination of several extensions (e.g. txt)
Size	TBD
Info about metadata (production and storage data, places) and standards	TBD
Data access and sharing	
Access to data	TBD
Reason for restricted access	TBD

Requirements to access	TBD
License for datasets	TBD
Data sharing	TBD
Reason for unavailable sharing	TBD
Archiving and preservation	
Data storage (including backup)	TBD
Period of archiving	TBD

4.6 Policies for access, sharing and reuse

Data generated during QU4LITY project will become confidential, ownership and management of intellectual property and access will be limited. For this purpose, policies for access, sharing and reuse have been stablished.

However, considering the description and instructions of the ORD pilot, QU4LITY consortium agrees on making publications and the research data related to the scientific publication open. More information about the open data available in section 4.8.3.1.

4.7 Partners Background

Partners have identified their background for the action (data, know-how or information generated before they acceded to the Agreement) which will be accessible to each other partners to implement their own tasks (under legal restrictions or limits previously defined). The partners should be able to access, mine, exploit, reproduce and disseminate the data. This should also help to validate the results presented in scientific publications. The partner's background acquired prior to the starting date of the project will remain the sole property of the originating partner, provided that it was presented in the CA.

4.8 Data Ownership and Access

In principle, the full dataset will be confidential and only the members of the consortium will have access on it. As described in GA, data generated are expected to be used internally as input by the other WPs. All the partners will have free access to the results generated during the project, the information needed for implementing their own tasks under the action and for exploiting their own results. Also, this

information will be available to EU institutions, bodies, offices or agencies, for developing, implementing or monitoring EU policies. However, such access rights are limited to non-commercial and non-competitive use.

- **4.8.1 Naming rules**

All data files will be saved using a standardised consistent file naming protocol agreed by the project partners, which will include relevant metadata to ensure their accessibility. The metadata standard proposed is the CERIF (Common European Research Information Format) however some datasets will follow different standards according to the internal policies of their owners.

- **4.8.2 Storage Information**

Data shared among project partners will be stored at the non-free access repository (newrepository.atosresearch.eu) created and maintained by ATOS, the project coordinator. The research data and documents generated during the project will remain at this repository for the whole project duration, as well as for 5 years after the end of the project. Finally, after the end of QU4LITY project the portal is going to be accommodated with other portals at the same server, so as to minimise the needed costs for its maintenance.

The datasets provided by partners for experimental purpose may be stored in different locations and kept for different period of time according to the policies and IPR of their owners.

- **4.8.3 Data sharing and dissemination**

Data will be reuse for corrective actions on the deployed strategies and actions will be suggested based on correlations by the automatic decision support mechanism. Research data results will be disseminated according to the Consortium Agreement in the form of conference, articles in a journal, specialist magazine/website outlets or conference proceedings for dissemination purposes.

- ***4.8.3.1 Open Access***

Open Access can be defined as the practice of providing online access to scientific information that is free of charge to the reader. In the context of R&D, open access typically focuses on access to "scientific information" or "research results", which refers to two main categories:

- Peer-reviewed scientific research articles (primarly published in academic journals)
- Research data

All patents and other publications will require prior agreement in respect both the content and the publication media. To this end each partner should notify the consortium members about the content and material they wish to publish/disseminate and a 30 days evaluation period will be provided. Every scientific

publication will be provided in open access mode, making it available to maximise access and reuse of the findings.

Regarding the research data only those related to the scientific publications will be open, after studying each dataset separately and confirming there are not good reasons to keep it closed. In addition, in some situations, the research data will be anonymised in order to preserve the confidentiality and know how.

Both, scientific information and open research data, will be available and published on a public repository linked to the project website. The data published in this repository will also made available on the OpenAIRE portal that is based on European OA repository infrastructure.

- ### 4.8.4 IPR management and security

As an innovation action close to the market, QU4LITY project covers high TRL technologies and aims at developing marketable solutions. The project consortium includes 33 partners from the private sector (Atos, Siemens, Philips Cl, Airbus, Whirlpool, Continental Automotive Gmbh, Nxtcontrol (Schnieder-Electric), +Gf+, Thyssenkrupp, Riastone, Kolektor, Danobat, Fagor Arrasate , Ikerlan, Prima Industrie, Visual Components, Guinea Hermanos, Engineering, Telefonica, Tttech, Mondragon Corporation, Synesis, Asti, Unparallel, Unimetrik, TTS, Atlantis, SQS, TXT, Sintef, Tubs-Iwf, ATB, Asociación Innovalia).

Those partners obviously have intellectual property rights on their technologies and data, on which their economic sustainability is at stake. Consequently, QU4LITY project consortium will protect data and crosscheck with the concerned partners before every data publication.

The share point is equipped with authentication mechanisms, so as to handle the identity of the persons and organisations that download them, as well as the purpose and the use of the downloaded dataset.

- ### 4.8.5 Data expire date

Copyright statements of the QU4LITY project will protect any written material produced during the project. As described in GA, the information supplied by any of the partners and the data and documents produced during the project will be protected for a period of five years after the project completion unless there are agreements between the partners.

The partners will keep for five years (after the end of the project) the original documents, digital and digitalised documents, records and other supporting documentation in order to prove the proper implementation of the action and the costs they declare as eligible.

5. Data management related to predictive maintenance

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 QU4LITY will be distributed freely.

6. Data management portal

On the one hand, the share point is implemented through a web-based platform which enables its user to easily access and effectively manage the various documents and datasets created throughout the development of the project.

Regarding the user authentication as well as the respective permissions and access rights, the following three user categories are foreseen:

- **Admin:** The admin has access to all of the datasets and the functionalities offered by the DMP and is able to determine and adjust the editing/access rights of the registered members and users (open access area). Finally, the Admin is able to access and extract the analytics concerning the visitors of the portal.
- **Member:** When someone successfully registers to the portal and is given access permission by the Admin, this person is then considered as a "registered member". All the registered members will have access to and be able to manage most of the collected datasets.

The share point will be easily and effectively managed by the members. A variety of graph, pie charts, etc is going to be employed for helping members to easily understand and elaborate the data. In particular, the architecture of the portal presents special interfaces organised to comply the information.

All tasks and datasets available in the DMP will be accompanied by a short description of the item.

On the other hand, QU4LITY website is an open website whose main objective is to disseminate the project purpose, strategies, activities and outcomes of the research. This communication tool is continuously updated with the project progress and dissemination activities/information.

The website has different areas: Project, partners, news&events, public deliverables and dissemination assets (this section will include articles and publications so as to be available for all the visitors) and contacts.

7. Conclusions

This report includes the Data management Plan (DMP) and describes the Research Data Information that is being generated during QU4LITY project as well as the challenges and constraints that need to be taken into account for managing it. In addition, it describes the updated procedures and the infrastructure implemented by the project to efficiently manage the produce data. The DMP is identified as starting point for the discussion with the community about the QU4LITY data management strategy.

In this version, a questionnaire has been designed to collect information on the already defined by the partners the data they are collecting and generating or they expect to produce and collect during the experimentation of the pilots. DMP is not a fixed document and it will be updated during the lifespan of the project each year. Thus, the DMP will be again and augmented with new datasets and results during project lifetime, submitting the final version in M36.

Regarding storage information, all the documents generated during the projects will be stored in the common repository point, provided by the coordinator which is the document management system of the project. This information, data and documents produced during the project will be protected for a period of a 5 years after the project completion, as it is described in GA.

8. Glossary

Participant Information Sheet

The information sheet is an important part of recruiting research participants. It ensures that the potential participants have sufficient information to make an informed decision about whether to take part in your research or not.

Consent Form

A form signed by a participant to confirm that he or she agrees to participate in the research and is aware of any risks that might be involved.

Metadata

Metadata is data that describes other data. Meta is a prefix that in most information technology usages means "an underlying definition or description." Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier.

Repository

A digital repository is a mechanism for managing and storing digital content. Repositories can be subject or institutional in their focus.

9. Bibliography

- Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26 July 2016: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 3.2, 25 March 2017: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- Webpage of European Commission regarding Open Access: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

List of tables

List of Abreviations

CA	Consortium Agreement
DB	DataBase
DEM	Dissemination and Exploitation Management
DMC	Data Matrix Code
DMP	Data Management Plan
EC	European Commission
FAIR	Findable, Accessible, Interoperable, and Reusable.
GA	Grant Agreement
IPR	Intellectual Property Rights
KPI	Key Performance Indicator
ORD	Open Research Data

Partners: