Z - B R E 4 K

Grant agreement nº: 768869
Call identifier: H2020-FOF-2017

Strategies and Predictive Maintenance models wrapped around physical systems for Zero-unexpected-Breakdowns and increased operating life of Factories

Z-BRE4K

Deliverable D8.4
Data Management Plan (DMP)

Work Package 8
Dissemination/Communication/Exploitation

Document type	:	Report
Version	:	3.0
Date of issue	:	02/06/2020
Dissemination level	:	PUBLIC
Lead beneficiary	:	TRIMEK

Executive Summary

Abstract	This is the third version of the Data Management Plan (DMP), a final update of the datasets and data flow that we are going to find in Z-BRE4K. This 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 will define which data will be collected and 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.
Keywords	Data Management Plan, Dataset, Open Data.

Revision history

Version	Author(s)	Changes	Date
1.0	Carlos Montoro García (TRIMEK), Silvia de la Maza (TRIMEK)	First draft	23/03/2018
1.1	Paloma Taboada (TRIMEK)	First update M18	01/03/2019
1.2	CORE, AIMEN and ATLANTIS	Contributions to M18 DMP	19/03/2019
1.3	ATLANTIS	Final review to M18 DMP	02/04/2019
1.4	CRIT	Contribution to M18 DMP	02/04/2019
2.0	Paloma Taboada (TRIMEK)	Final M18 version	02/04/2019
2.1	Pablo Salvo (INNOVALIA) & Silvia de la Maza (TRIMEK)	Updates for the 3rd version (M30)	15/05/2020
2.9	Silvia de la Maza	Final M30 version	22/05/2020
3.0	AIMEN, CRIT, INNOVALIA & TRIMEK	Peer reviewed & Submitted version	28/05/2020

Table of contents

1 ABBREVIATIONS

Abbreviation	Name
DSS	Decision Support System
FMECA	Failure Modes, Causes and Effects Analysis
DMP	Data Management Plan
ORD Pilot	Open Research Data Pilot
EU	European Union
EC	European Commission
OCB	Orion Context Broker
IPR	Intellectual Property Rights
WP	Work Package
GA	Document of Grant Agreement
DEM	Dissemination and Exploitation Manager
QIF	Quality Information Framework
CCF	Common-cause failures
FM	Failure Mode
RUL	Remaining Useful Life
KRI	Key Risk Indicators
RCA	Root Cause Analysis
IDS	Industrial Data Space
CMM	Coordinate Measuring Machine

2 ACKNOWLEDGMENTS

3 SUMMARY

The present report is focused on the preparation of the Data Management Plan (DMP) for Z-BRE4K project, an updated version in which the datasets that are being generated, as well as the data flow on the Z-BRE4K solutions finally proposed, can now be presented in a more approximate way to the final approach of the project. The objective of the DMP is to provide 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.

The DMP of Z-BRE4K 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, promoting an update and expansion of the detail of the datasets and results as we advance throughout the life of the Z-BRE4K project, as it can be seen on this 3rd version of the DMP.

This document is structured in the following sections:

- It starts with an introduction to the purpose of the creation of a DMP and the importance of generating open data indicating our participation in the pilot on open research data.
- Then we will focus on the development of Z-Bre4k DMP, describing the methodology and the activities carried out to perform it, followed by the metadata and datasets list description.
- On this same section we will continue with the in-deep description of each dataset following a structured based on the datasets generated per task and partner.
- Next will be the policies for access, data sharing and data ownership summary description, including the portals description where the Z-Bre4k is shared and can be found.
- Finally, some conclusions will be provided giving a brief summary of the Data Management Plan evolution, same way as datasets did.

4 INTRODUCTION

The amount of data generated is continuously increasing. whit the generation of new data and the re-use of previously generated data is intended to obtain exploitable results for the project purpose but also to derive new scientific findings. All this information could only be useful if a data management plan, as a methodology, is followed. If the generated datasets are well documented according to accepted and trusted standards which enable the recognition of suitable data by negotiated agreements on standards, quality level and sharing practices. For this purpose, the DMP defines the strategies to preserve and store the data over a defined period of time in order to ensure their availability and re-usability after the end of Z-BRE4K project.

In the case of Z-BRE4K project, a variety of datasets will be generated with the overall objective to develop predictive maintenance manufacturing strategies in order to increase the operating life of production systems. Z-BRE4K aims at contributing to the eradication of machine and component failures in manufacturing, providing a sustainable manufacturing system with high quality products, reducing resource consumption and waste generation. Thus, research activities are focused on obtaining data from the machines and field devices, in order to develop a complete predictive maintenance system.

As it has been mentioned before, Z-BRE4K participates in a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot), that aims improving and maximizing access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions.

Following the recommendation of the EC and the ORD pilot guidelines, the DMP is considered as a living document (D8.4) and for that reason, few versions with updates and extensions of the datasets and results have been elaborated on the scheduled milestones (M6, M18 & M30), according to the progress of activities of Z-BRE4K project. The DPM of Z-BRE4K 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).

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.

4.1 Building a DMP in the context of H2020

The EC provided 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 Z-BRE4K 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 surveys 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 the 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 Z-BRE4K 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.

5 DATA MANAGEMENT PLAN (DMP)

5.1 General description

This document consists of the third version of the project's DMP, the third version of the D8.4. As it has been explained previously, the main purpose of a DMP is to provide an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project.

In this document we will find described the Research Data with the metadata attached, an 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 reuse. In addition, DMP of Z-BRE4K project reflects the current status of discussion within the consortium about the data that will be produced.

5.2 Activities of Data Management Plan

The DMP is a dynamic document, updated throughout the whole project lifecycle. Now we find ourselves on the final version of this report, delivered before the end of the project (M32), reflecting the changes produced in the generation of datasets and the dataflow as the general and particular idea of each part of the project evolves, describing some aspects that weren't detailed on previous versions of the DMP regarding ownership, utility, the plans implemented for sustainable storage and accessibility of the data, even beyond the project's duration. Also, some lessons learnt will be reflected on the conclusions.

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 Z-BRE4K concerning the data management are planned as follows:

- A Knowledge manager role to be led by the Dissemination and Exploitation Manager (DEM).
- 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.

5.3 Register on numerical data sets generated or collected in Z-BRE4K

The intention of the DMP is to describe numerical model or observation datasets collected or created by Z-BRE4K activities during the runtime of the project. The register on numerical data sets has to be understood as living document, which will be updated regularly during project lifetime.

The operational phase of the project started in January 2018, so there was no dataset generated or collected before the first delivery date of this DMP (Month 6). Further on following versions, datasets have been registered and tracked through the elaboration of various questionnaires. That is, this is not a fixed document, it has been updated and augmented with new datasets and results along the duration of Z-BRE4K project.

According to the Guidelines of Open Research Data in Horizon 2020, in order to define how the Z-Bre4k project data will be managed and shared by project partners, as a matter of making research data findable, accessible, interoperable and re-usable, the first approach of the DMP was performed following the questionnaire included above.

<DATASET_NAME><PARTNER_NAME>	
Data Identification	
Data set description	
Source (e.g. which device?)	
Partners activities and responsibilities	
Partner owner of the device	
Partner in charge of the data collection (if different)	
Partner in charge of the data analysis (if different)	
Partner in charge of the data storage (if different)	
WPs and tasks	
Standards	
Info about metadata (Production and storage dates, places) and documentation?	
Standards, Format, Estimated volume of data.	
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	
Data sharing, re-use and distribution (How?)	
Embargo periods (if any)	

Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	

This is a description of the detailed information that should be filled in the previous questionnaire 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.

In this last version of the document, as a matter to make more visual the dataflow between the different WPs and elements that compound the global Z-BRE4K solution, a new questionnaire has been promoted to be filled by every partner, covering some other aspects that maybe weren't so detailed on the previous one. In any case, the intention was to achieve that goal as a living document to update the dataset status of the whole project.

This mentioned second questionnaire that aims at identifying and gathering all the datasets to be generated by all consortium partners during the lifetime project, covers the following aspects:

1) Scope

o State the purpose of the data generation/collection
o Explain the relation to the objectives of the project/WP/Task

2) Types

o Are the data digital/hard copies or both?
o What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)
o Is the data generated or collected from other sources under certain terms and conditions?
o How is generated/collected? Specify the origin of the data and instruments/tools that will be used.
o State the expected size of the data (if known)
o Standards

3) Ownership

o Is another organization contributing to the data development?

4) Reuse of existing data

o Specify if existing data is being re-used (if any)

5) Data use

o How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful

6) Dissemination Level of Data

o Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?

7) Storage and disposal

o How will this data be stored?
o How long is it required to keep the data? Expire date. Will revisions be kept?

8) Rules for Data naming

o Standards or internal rules

5.4   Metadata for Data Management

An initial plan of research data has been explored in Annex 1 of the GA. The datasets list is provided in the table below, while the nature and details for each dataset are presented in the next section. The initial list has been updated with new research data identified throughout project development.

Table 1. Research data that will be collected and generated during Z-BRE4K

Research Data	Partners
Data structures with machine system signatures (healthy and deteriorated conditions)	ATLANTIS
Machine Systems deterioration thresholds for predicting failures	ATLANTIS
RCA data structures for identifying the root cause of a failure mode (FM)	ATLANTIS
Common-cause failures (CCF) and the FMs with most severe effects; Data structures of a Criticality Matrix; KRIs	ATLANTIS
Thresholds for trends forecasting in relation to condition monitoring and FM prediction	ALL PARTNERS
Failure mode detection/prediction efficiency data: false alarm rate, precision, recall, F-Measure	ALL PARTNERS
Failure prediction efficiency data: positive prediction rate	ALL PARTNERS
Failure Modes and data structures with associated condition monitoring parameters, as well as ambient environment parameters (usage, environmental conditions, materials, quality of resources, etc).	ALL PARTNERS
Data structures and method for retrofitting physically based models of machinery systems	FRAUNHOFER
Data from the comparative assessment (i.e. with and without Z-BRE4K) in the 3 use cases: difference in efficiency, effectiveness, production planning and schedule, failure predicted/prevented, remedy actions	SACMI-CDS, PHILIPS, GESTAMP
Discrete Event Modelling – cost function generation to optimize prediction and failure analysis	BRUNEL
Improved functionalities of i-LiKe knowledge management and DSS suite	HOLONIX
Condition Monitoring data framework for CMM predictive maintenance strategies	TRIMEK
Quality control for process predictive maintenance strategies	TRIMEK
Production data for predictive maintenance	CORE

As aforementioned, partners have characterized their research data and associated software and/or used in the project whether these are:

- Findable / Discoverable: can be discovered by means of an identification mechanism such as Digital Object Identifier.
- Accessible: define in what modalities, the scope of the action, stablish the licenses and define the IPR.
- Assessable and intelligible: allow to third parties to make assessments.
- 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).
- Interoperable to specific quality standards allow data exchange between researchers, institutions, organizations countries, re-combinations with different datasets, data exchange, compliant with available software applications.

5.5   Data description

As it has been described on section4.3, In this final version of the DMP, we have elaborated a different questionnaire to be completed by all the partners in order to cover some other aspects that weren't so detailed with the previous one, but also to re-structure the dataset connections providing a distinct point of view of the dataflow.

This template includes the following information items:

1) Scope

2) Types

3) Ownership

4) Reuse of existing data

5) Data use

6) Dissemination Level of Data

7) Storage and disposal

8) Rules for Data naming

Another novelty with respect to the previous deliverables, is that in this case, the datasets will be ordered by tasks and work packages.

5.5.1 Datasets per Task and WP

All partners have identified the data that will be produced in the different project activities.

WP2 - Operating system and networked machine simulators

Task 2.2 - Industrial Data Spaces (IDS), connectors and containers

Identification information	
Author:	INOVA+ (P11)
WP:	2
WP Leader:	INNOVALIA
Task:	T2.2
Scope	
Scope: State the purpose of the data generation/ collection	The IDS Connector collects and translates proprietary data (from the end-user and shopfloor level) to make it available at the Context Broker for consume by the other modules in the platform.
Scope: Explain the relation to the objectives of the project/WP/Task	the IDS Connectors is part of the data flow path, i.e., it is required to fulfil the overall objective on integrating all components into one single platform.
Data Type	
Are there data digital/hard copies or both?	digital
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	Data collected has 3 different formats at its "source" (*.txt, *.JSON and *.csv). It is then translated into NGSI format (kind of a modified JSON).
Is the data generated or collected from other sources under certain terms and conditions?	The data is collected at pre-defined frequencies, formats and purposes, thus yes.
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Data is/will be collected directly form a proprietary database. A pull mechanism is implemented in the Connector to retrieve the source data at prescribed rates.
State the expected size of the data (if known)	
Standards	JSON, NGSI
Ownership	
Owner	End Users
Is another organization contributing to the data development?	raw data provided by end-user. data filtering rules provided by CORE and ATLANTIS
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	not used for other purposes than the Z-BRE4K processing
How will this data be exploited and/or shared/made accessible for verification and	n/a

re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	since data is not anonymized, it is crucial to be kept confidential. Data is proprietary and private.
Storage & Disposal	
How will this data be stored?	temporary storage in Context Broker + Permanent storage in offline database
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	rules yet to be defined.
Rules of Data Naming	
Standards or internal rules	n/a
Release Date	
Can data be uploaded in an open research data pilot? When?	need to check with end-user (owner)

Task 2.4 - SoN and interfacing with inspection & metrology infrastructure

Identification information	
Author:	TRIMEK (9)
WP:	2
WP Leader:	INNOVALIA
Task:	T2.4
Scope	
Scope: State the purpose of the data generation/ collection	Data coming from the M3 Gage component used for the condition monitoring and status verification of the CMM in order to ensure the adequate performance and to avoid unexpected breakdowns.
Scope: Explain the relation to the objectives of the project/WP/Task	Data collection through the M3Box (T2.4) from the CMM operative condition regarding demonstration activity on the GESTAMP Use case (T6.2).
Data Type	
Are there data digital/hard copies or both?	Both
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for	Log files (.TXT), .CSV and .M3 file extensions

example .xls files, .ppt files, emails, .doc files)	
Is the data generated or collected from other sources under certain terms and conditions?	M3 Box gathers the data from the CMM controller, regarding the electromechanical sensor data; and also, from M3 software (tetracheck), regarding the status verification operation.
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Data gathered, processed and stored by M3 BOX and M3 Analytics module.
State the expected size of the data (if known)	~50 Kb per file (electrical current monitoring / CMM status verification)
Standards	OPC-UA
Ownership	
Owner	TRIMEK
Is another organization contributing to the data development?	INNOVALIA
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	This data will be exploited by Z-BRE4K consortium. It will be useful mainly for TRIMEK and its customers.
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	The full dataset will be confidential, but the reports and semantic model will be public.
Storage & Disposal	
How will this data be stored?	M3 Cloud
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	**TBD**
Rules of Data Naming	
Standards or internal rules	
Release Date	
Can data be uploaded in an open research data pilot? When?	CMM status verification data comes from an own software performance, so it is confidential data.

Task 2.5 - CPS-based condition monitoring and embedded processing

Identification information	
Author:	AIMEN (P1)
WP:	2
WP Leader:	INNOVALIA
Task:	T2.5
Scope	
Scope: State the purpose of the data generation/ collection	The data collected will be used to feed the cognitive monitoring embedded system to reduce its dimensionality extracting the most important features. There also will be data collected previously to the integration of the system in order to adjust the system to its purpose. The data generated will be the result of the feature extraction process and it will be useful to speed up communications, optimize storage and help the predictive systems.
Scope: Explain the relation to the objectives of the project/WP/Task	The main objective of the project is to avoid the failures in industrial processes and one of the most common problems that predictive or corrective systems have to do it is the huge amount of data generated. The cognitive monitoring embedded system will be able to reduce this amount of data maintaining the most useful information.
Data Type	
Are there data digital/hard copies or both?	There will be both types. The data used to adjust the initial system will be a hard copy of the data collected in the process. The data used to generate the output of the system will be digital data acquired from the process sensors.
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	The data collected format depends on the sensor or system the data will be acquired for. The data generated will be in JSON or HDF5 formats, usually used to store scientific data, but it will depend on demand.
Is the data generated or collected from other sources under certain terms and conditions?	N/A

How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	The data will be collected from all the sensors of the process, the tools used will be middleware software to acquire the data or to communicate with the system connected to the sensors. The generated data will be the result of the feature extraction done with machine learning algorithms and/or expert systems.
State the expected size of the data (if known)	N/A
Standards	N/A
Ownership	
Owner	
Is another organization contributing to the data development?	The end users are the partners that provide the data.
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	The data collected by the system could be reused to retrain them in order to improve its performance
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	N/A
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	The data collected by the cognitive monitoring system will be accessible only to AIMEN Technology Centre. The data provided by the system will be also accessible to the technology partners of ZBRE4K consortium that develop components to predict failures in the process.
Storage & Disposal	
How will this data be stored?	The data generated by the cognitive monitoring embedded system will be stored in the main ZBRE4K storage system using preferably HDF5 format.
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	N/A

Rules of Data Naming	
Standards or internal rules	N/A
Release Date	
Can data be uploaded in an open research data pilot? When?	

WP3 - Knowledge and Predictive Modelling

Task 3.1 - Semantic modelling of production assets, products and processes

Identification information	
Author:	EPFL (P13)
WP:	WP3
WP Leader:	BRUNEL
Task:	Task 3.1
Scope	
Scope: State the purpose of the data generation/ collection	The reference model for annotation and description of knowledge to represent manufacturing system performance in the context of Z-BRE4K
Scope: Explain the relation to the objectives of the project/WP/Task	The result of Task 3.1. It is implemented under Task 3.2
Data Type	
Are there data digital/hard copies or both?	Digital
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	RDF
Is the data generated or collected from other sources under certain terms and conditions?	n/a
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Design methodology created by EPFL
State the expected size of the data (if known)	<1MB
Standards	RDF, OWL
Ownership	
Owner	EPFL

Is another organization contributing to the data development?	n/a
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	n/a
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	The reference model is published. Anyone can access/reuse it.
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Published
Storage & Disposal	
How will this data be stored?	n/a
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	n/a
Rules of Data Naming	
Standards or internal rules	n/a
Release Date	
Can data be uploaded in an open research data pilot? When?	To be defined

Task 3.2 - Knowledge Base System (KBS) with asset/product/process signatures

Identification information	
Author:	HOLONIX (P12)
WP:	3
WP Leader:	BRUNEL
Task:	T3.2
Scope	
Scope: State the purpose of the data generation/ collection	The Condition Monitoring component based on HOLONIX I-like machine platform collect/use data by third parties to generate a set of machine KPIs, in order to monitoring the machine status
Scope: Explain the relation to the objectives of the project/WP/Task	Data collected by the Condition Monitoring module are used to monitor machines/systems behaviour to enable a fast reaction in case of incipient failures.

	Moreover, the integration with predictive algorithms that will be carried out during the project will help in estimate the residual life of the machine/component. This concurs to the overall objective of Z-BRE4K project to obtain a zero-breakdown manufacturing.
Data Type	
Are there data digital/hard copies or both?	Digital. Data collected and/or generated are stored within several relational and no-relational DBMSes. The DBMSes are installed in virtual or physical servers owned by HOLONIX
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	The Condition Monitoring component based on HOLONIX I-like machine platform collects data that comes from machines in several formats depending on specific case through AUTOWARE middleware. Data exchange will be most likely in JSON format. The Condition Monitoring component generates KPIs values that will be visualized by user in graphical, table or textual format on the User Interface.
Is the data generated or collected from other sources under certain terms and conditions?	In the Z-BRE4K specific context the data exchange will be managed by AUTOWARE solutions
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	From condition monitoring module viewpoint, the origin of data is the AUTOWARE IDS data repository. In a wider perspective, data comes from machines and sensors installed nearby. The methodology of data collection from AUTOWARE solutions has still to be defined.
State the expected size of the data (if known)	About 100 KB per minute
Standards	O-DF (Open data Format)
Ownership	
Owner	End Users
Is another organization contributing to the data development?	End users will be the provider of raw data, they will be stored in the AUTOWARE IDS repository before reaching the condition monitoring module. Here they will be

	treated and reorganised in KPIs. Algorithm to predict the component behaviour with respect to failure developed within the consortium will be integrated in the I-LiKe machine platform.
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	HOLONIX will utilize data generated and collected only within the Z-BRE4K context
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	First, the data will be made available only for the consortium members. In a second stage, and depending on the agreement within the consortium, it can be considered to extend the data availability. This kind of data can be useful for machine owner to monitor the machine status and plan maintenance interventions, and to machine builders to improve the machine reliability.
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	In a first stage, only member of Z-BRE4K consortium will have access to data collected and generated by condition monitoring module within the Z-BRE4K context. It can be evaluated within the consortium member and, in particular, with the end users the possibility to disclose the data.
Storage & Disposal	
How will this data be stored?	The data will be stored within several relational and no-relational DBMSes installed in virtual or physical servers owned by HOLONIX
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	The overall duration will be defined within Z-BRE4K project. For sure all the project duration and after per 5 years.
Rules of Data Naming	
Standards or internal rules	i-Like machine platform has its own case-dependant naming strategies.
Release Date	
Can data be uploaded in an open research data pilot? When?	Data from CDS are available since M18

Task 3.3 - FMECA and KRI models and criticality analysis tools

Identification information	
Author:	ATLANTIS (P2)
WP:	3
WP Leader:	BRUNEL
Task:	T3.3
Scope	
Scope: State the purpose of the data generation/ collection	FMECA component
Scope: Explain the relation to the objectives of the project/WP/Task	Models, rules, data for FMECA component
Data Type	
Are there data digital/hard copies or both?	Digital
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	CSV JSON Binary PDF Text
Is the data generated or collected from other sources under certain terms and conditions?	Collected from the shop floor (sensors, cameras, actuators)
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	From signals From user input Batch files
State the expected size of the data (if known)	Several GBs per day
Standards	JSON, XML
Ownership	
Owner	To be described
Is another organization contributing to the data development?	Only partners of the consortium
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	n/a
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	Used for modelling
Dissemination Level of Data	

Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Avoiding storage of sensitive data. Stored encrypted always. Limited access to confidentiality data
Storage & Disposal	
How will this data be stored?	Z-BRE4K's repository. Component repository (if needed).
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	n/a
Rules of Data Naming	
Standards or internal rules	n/a
Release Date	
Can data be uploaded in an open research data pilot? When?	n/a

Task 3.4 - *Machine Learning System and predictive, preventive, diagnostic analytics*

Identification information	
Author:	ATLANTIS (P2)
WP:	3
WP Leader:	BRUNEL
Task:	T3.4
Scope	
Scope: State the purpose of the data generation/ collection	Fault Prediction/Detection
Scope: Explain the relation to the objectives of the project/WP/Task	Data will used as input to fault detection and prediction tools
Data Type	
Are there data digital/hard copies or both?	Digital
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	JSON
Is the data generated or collected from other sources under certain terms and conditions?	Collected from INOVA+ and transferred through their IDS connector
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Collected and processed by INOVA+, transferred using their IDS connector
State the expected size of the data (if known)	Several GBs per day
Standards	JSON
Ownership	
Owner	PHILIPS

Is another organization contributing to the data development?	Only partners of the consortium
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	Historical data are sued to train ML prediction models
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	Used by the ATLANTIS Engineering to detect/predict faults and to fine tune the detection tools and to train the prediction ones
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Anonymized data. Limited access to confidentiality data
Storage & Disposal	
How will this data be stored?	Z-BRE4K's repository. Component repository (if needed).
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	n/a
Rules of Data Naming	
Standards or internal rules	n/a
Release Date	
Can data be uploaded in an open research data pilot? When?	n/a

Identification information	
Author:	BRUNEL
WP:	3
WP Leader:	BRUNEL
Task:	T3.4
Scope	
Scope: State the purpose of the data generation/ collection	Prediction Failure DSS FMECA component
Scope: Explain the relation to the objectives of the project/WP/Task	Data will used to model prediction models to generate RUL and probability of breakdown.
Data Type	
Are there data digital/hard copies or both?	Both
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for	CSV JSON Binary

example .xls files, .ppt files, emails, .doc files)	PDF Text Maintenance reports etc.
Is the data generated or collected from other sources under certain terms and conditions?	Collected from the shop floor (sensors, cameras, actuators)
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	From signals From user input Batch files web services etc.
State the expected size of the data (if known)	Several GBs per day
Standards	Not known at this stage
Ownership	
Owner	
Is another organization contributing to the data development?	Only partners of the consortium
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data	n/a
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use	the outcomes of predictions will be used by DSS for maintenance suggestions.
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Avoiding storage of sensitive data. Stored encrypted always. Limited access to confidentiality data
Storage & Disposal	
How will this data be stored?	Z-BRE4K's repository. Each component's repository (if needed).
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions	n/a
Rules of Data Naming	
Standards or internal rules	n/a
Release Date	
Can data be uploaded in an open research data pilot? When?	n/a

Identification information	
Author:	CORE (P4)

WP:	3
WP Leader:	BRUNEL
Task:	3.4

Scope	
Scope: State the purpose of the data generation/ collection	Training Machine learning models
Scope: Explain the relation to the objectives of the project/WP/Task	Provide tools to predict / detect degradation of machines
Data Type	
Are there data digital/hard copies or both?	digital
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	Csv, json, db files
Is the data generated or collected from other sources under certain terms and conditions?	Collected from HOLONIX Web Services and Orion Brokers
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	End Users' sensors, data acquisition systems, communication middleware by HOLONIX, INOVA
State the expected size of the data (if known)	NA
Standards	
Ownership	
Owner	SACMI, CDS, PHILIPS
Is another organization contributing to the data development?	No
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	NA
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	Model training, Model validation, Exploration Analysis
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Confidential
Storage & Disposal	
How will this data be stored?	Digital format

How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	At least a year for validation purposes
Rules of Data Naming	
Standards or internal rules	NA
Release Date	
Can data be uploaded in an open research data pilot? When?	No, confidential data

Identification information	
Author:	IMEC (15)
WP:	3
WP Leader:	BRUNEL
Task:	3.4
Scope	
Scope: State the purpose of the data generation/ collection	Develop algorithms for fault prediction and remaining useful life estimation. Our tasks make use of SACMI and PHILIPS data, we do not generate data.
Scope: Explain the relation to the objectives of the project/WP/Task	The algorithms can be used for diagnosis and predictive maintenance.
Data Type	
Are there data digital/hard copies or both?	Digital
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	.json .xlsx .txt .csv
Is the data generated or collected from other sources under certain terms and conditions?	-
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Sensors and stored in the cloud by PHILIPS/SACMI/HOLONIX
State the expected size of the data (if known)	Unknown
Standards	-
Ownership	
Owner	PHILIPS/SACMI
Is another organization contributing to the data development?	-
Data Use & Utility	

Specify if existing data is being re-used (if any). Specify the re-use of data …	-
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	-
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	According to Philips and SACMI
Storage & Disposal	
How will this data be stored?	According to HOLONIX
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	-
Rules of Data Naming	
Standards or internal rules	-
Release Date	
Can data be uploaded in an open research data pilot? When?	-

WP4 - Design of Strategies and Integration of Intelligence

Task 4.2 Decision support for Predictive & JIT maintenance towards operational optimisation

Identification information	
Author:	ATLANTIS (P2)
WP:	4
WP Leader:	EPFL
Task:	T4.2
Scope	
Scope: State the purpose of the data generation/ collection	DSS
Scope: Explain the relation to the objectives of the project/WP/Task	Data will used to model DSS
Data Type	
Are there data digital/hard copies or both?	Digital
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for	CSV JSON Binary

example .xls files, .ppt files, emails, .doc files)	PDF Text
Is the data generated or collected from other sources under certain terms and conditions?	Collected from the shop floor (sensors, cameras, actuators)
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	From signals From user input Batch files
State the expected size of the data (if known)	Several GBs per day
Standards	JSON, XML
Ownership	
Owner	To be described
Is another organization contributing to the data development?	Only partners of the consortium
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	n/a
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	To visualise processes Create visual KPI's
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Avoiding storage of sensitive data. Stored encrypted always. Limited access to confidentiality data
Storage & Disposal	
How will this data be stored?	Z-BRE4K's repository. Component repository (if needed).
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	n/a
Rules of Data Naming	
Standards or internal rules	n/a
Release Date	
Can data be uploaded in an open research data pilot? When?	n/a

TASK 4.4 - Decision support for Predictive & JIT maintenance towards operational optimization

Identification information

Author:	INNOVALIA (P5)
WP:	4
WP Leader:	EPFL
Task:	T4.4

Scope	
Scope: State the purpose of the data generation/ collection	Collect quality control and CMM condition monitoring data from M3Gage and transform/contextualize it into NGSI for its publication on the Orion Context Broker (OCB).
Scope: Explain the relation to the objectives of the project/WP/Task	M3 BOX acts as a middleware (T4.4) in order to share M3Gage data with the other components and Z-BRE4K services.

Data Type	
Are there data digital/hard copies or both?	Both
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	The data collected follows these formats: .QIF, .DMO, .CSV .TXT, .M3 file extensions but also XML files.
Is the data generated or collected from other sources under certain terms and conditions?	From the M3 Gage
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	The data collected comes from the M3Gage regarding Quality control data and condition monitoring of the CMM status data.
State the expected size of the data (if known)	*Look for TRIMEK datasets*
Standards	QIF, OPC-UA

Ownership	
Owner	TRIMEK/GESTAMP
Is another organization contributing to the data development?	TRIMEK

Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	No
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	This data will be exploited by Z-BRE4Kconsortium. It will be useful to demonstrate the IDS connector function.

Dissemination Level of Data	

Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	The full dataset will be confidential, but the reports and semantic model will be public.
Storage & Disposal	
How will this data be stored?	The data will be stored on M3 Cloud and shared through the data space
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	**TBD**
Rules of Data Naming	
Standards or internal rules	The data will be accessible to Z-BRE4K consortium members including the commission services.
Release Date	
Can data be uploaded in an open research data pilot? When?	No as it is confidential

WP6 - Demonstration Activities

Task 6.1 - DEMO #1 –CDS-SACMI

Identification information	
Author:	SACMI (8)
WP:	6
WP Leader:	SACMI
Task:	6.1
Scope	
Scope: State the purpose of the data generation/ collection	Demonstration activities
Scope: Explain the relation to the objectives of the project/WP/Task	Data collection from pilot demonstrations: technical, user acceptance and factory impact indicators. Demonstration activities and the data originated by them are related to the WP6 and mainly T6.1 and 6.5.
Data Type	
Are there data digital/hard copies or both?	Both
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	For the reports: .xls files, .ppt files and .doc files. - Rtf documents for user acceptance questionnaires/online instruments, spreadsheets with data for technical KPIs.

	- Dedicated User Manuals and documentation for technical partners, including sensor, alarms and FMECA information. - Alarm reports (.xml/.csv). - Maintenance actions (.xml/.csv). - Condition Monitoring Data. - Periodic Predictive Maintenance information (anomaly detection, RUL estimation, Failure Prescription). - Predictive Maintenance reports, notifications (i.e.: automatic e-mails, sms, push-up notifications).
Is the data generated or collected from other sources under certain terms and conditions?	N.A.
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Alarms and condition monitoring are generated based on SACMI's automation platform (PLC + sensors) and HOLONIX i-Like Machines, based on KBS Maintenance report information is generated by CDS personnel through i-Like Machines interface PdM information is generated by PdM technical partners running available data/information on their modules Notifications may be generated by User interfaces (ILM – HOLONIX and DSS – ATLANTIS).
State the expected size of the data (if known)	1-5 GB/year and machine
Standards	Unknown
Ownership	
Owner	Condition monitoring data, alarms and maintenance reports belongs to CDS; Whereas FMECA and design & mechanical engineering information belongs to SACMI
Is another organization contributing to the data development?	Yes. Historical data collected in the previous years. Technical Partners involved in SACMI-CDS use case
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	N.A.

How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	Data is confidential
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Sensitive data only for SACMI.
Storage & Disposal	
How will this data be stored?	Unknown
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	At least 10 years.
Rules of Data Naming	
Standards or internal rules	Unknown
Release Date	
Can data be uploaded in an open research data pilot? When?	No. Partially, some R&D outputs not involving SACMI and CDS core information may be released (i.e.: ML algorithm applied for the RTF activities, e.g. regression based on condition monitoring, failure modes and health-based thresholds) provided that no confidential information is involved.

Task 6.2 - DEMO #2 – GESTAMP CHASSIS

Identification information	
Author:	TRIMEK (9)
WP:	6
WP Leader:	SACMI
Task:	T6.2
Scope	
Scope: State the purpose of the data generation/ collection	Data coming from the M3 Gage component used for the digitalization of a physical object, mainly for the dimensional inspection, this is, the quality control through a Point-of-Clouds performance.
Scope: Explain the relation to the objectives of the project/WP/Task	Data collection through the M3Gage regarding the quality control of the parts as

	one of the crucial steps of the GESTAMP production use case (T6.2).
Data Type	
Are there data digital/hard copies or both?	Both
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	.QIF, .DMO, .CSV and .M3 file extensions
Is the data generated or collected from other sources under certain terms and conditions?	Data is gathered through the M3 Gage
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Data gathered and processed by M3 Gage and stored on the M3 Cloud.
State the expected size of the data (if known)	~1Mb each part measured
Standards	QIF
Ownership	
Owner	GESTAMP/TRIMEK
Is another organization contributing to the data development?	
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	This data will be exploited by Z-BRE4K consortium.
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	The full dataset will be confidential, but the reports and semantic model will be public.
Storage & Disposal	
How will this data be stored?	M3 Cloud
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	**TBD**
Rules of Data Naming	

Standards or internal rules	The data will be accessible to Z-BRE4K consortium members including the commission services.
Release Date	
Can data be uploaded in an open research data pilot? When?	No, is confidential data.

Task 6.3 - DEMO #3 – PHILIPS

Identification information	
Author:	PHILIPS (P14)
WP:	6
WP Leader:	SACMI
Task:	T6.3
Scope	
Scope: State the purpose of the data generation/ collection	Create feedback on demonstrator within PHILIPS
Scope: Explain the relation to the objectives of the project/WP/Task	Create feedback on demonstrator within PHILIPS
Data Type	
Are there data digital/hard copies or both?	Both
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	Some reports and presentations on KPI's and some data files on performance of machines.
Is the data generated or collected from other sources under certain terms and conditions?	No
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Data is collected by PHILIPS data system. This collects data from machines and products. It generates reports on KPI's etc.
State the expected size of the data (if known)	Unknown
Standards	Unknown
Ownership	
Owner	PHILIPS
Is another organization contributing to the data development?	Only Z-BRE4K and PHILIPS will contribute.
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	Existing data will be reused for comparison.

How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	Data is confidential. So only accessible for Z-BRE4K consortium.
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	In a first stage, data will be available only for members of Z-BRE4K consortium. A possible disclosure must be agreed with the consortium member. Data will be shared outside Z-BRE4K consortium only if explicitly agreed by the consortium.
Storage & Disposal	
How will this data be stored?	PHILIPS servers
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	Yes
Rules of Data Naming	
Standards or internal rules	Unknown
Release Date	
Can data be uploaded in an open research data pilot? When?	No

Task 6.5 - Data collection from Demonstrations and Evaluation

Identification information	
Author:	CRIT (7)
WP:	6
WP Leader:	SACMI
Task:	6.5
Scope	
Scope: State the purpose of the data generation/ collection	Collect feedback on demonstration activities within the three use cases of the project: SACMI, PHILIPS and GESTAMP
Scope: Explain the relation to the objectives of the project/WP/Task	Feedback collection on demonstration activities are related to the WP6 and mainly T 6.5
Data Type	
Are there data digital/hard copies or both?	Both
What types of data will the WP generate/collect? Specify the types and	Primarily doc files and .xls files.

formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	Rtf documents for user acceptance questionnaires/online instruments, data of ERP-MES-SCADA repositories, spreadsheets with data for technical KPIs, reports and presentations on KPI's.
Is the data generated or collected from other sources under certain terms and conditions?	Some data may also be collected from other customers.
How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Data collection procedure will be definitively established around Month 28. It is expected to use each end user system for data collection and generation and to use also users' acceptance questionnaires, impact check lists and data collection forms.
State the expected size of the data (if known)	Unknown
Standards	Unknown
Ownership	
Owner	End users
Is another organization contributing to the data development?	Not expected, only Z-BRE4K and use cases will contribute
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	Yes, some historical data will be reused as starting point for the project as well as comparison.
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	Reports for the deliverables, descriptions of the activities performed within the three use cases, stakeholders involved, main findings, input for exploitation material will be public. Trend results will be shared but numeric and detailed data will not be available.
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Data are expected to be confidential, only accessible for Z-BRE4K consortium. These data will be helpful for preparing the deliverables of the project.
Storage & Disposal	
How will this data be stored?	Deliverable 6.5 (Confidential), public data within T6.5 may be shared through Zenodo,

	Github or other repositories for further research
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	Unknown
Rules of Data Naming	
Standards or internal rules	Unknown
Release Date	
Can data be uploaded in an open research data pilot? When?	No, in case public data is generated, it will be released together with D6.5 (by M42)

WP7 - Valorization and Market Assessment

T7.3 - Roadmapping for scale-up, uptake and market replication of Z-BRE4K

Identification information	
Author:	CRIT (7)
WP:	7
WP Leader:	INNOVA
Task:	7.3
Scope	
Scope: State the purpose of the data generation/ collection	Collect feedback on the diffusion and interest towards Predictive Maintenance Technologies by interviewing manufacturing companies for delivering a roadmap towards the scale up, uptake and market replication of Z-BRE4K.
Scope: Explain the relation to the objectives of the project/WP/Task	Feedback collection on market and customer segments, coming from organisations external to Z-BRE4K Consortium (industrial stakeholders, prospect customers)
Data Type	
Are there data digital/hard copies or both?	Both
What types of data will the WP generate/collect? Specify the types and formats of data generated/collected (for example .xls files, .ppt files, emails, .doc files)	Primarily doc files and .xls files. Report, graphs
Is the data generated or collected from other sources under certain terms and conditions?	N/A

How is generated/collected? Specify the origin of the data and instruments/tools that will be used.	Data are collected via Microsoft Forms or paper questionnaire and Word documents
State the expected size of the data (if known)	Unknown
Standards	Unknown
Ownership	
Owner	Z-BRE4K consortium, lead partner: CRIT
Is another organization contributing to the data development?	CORE, ATLANTIS, HOLONIX, AIMEN, AIC, INOVA, INNOVALIA
Data Use & Utility	
Specify if existing data is being re-used (if any). Specify the re-use of data …	No
How will this data be exploited and/or shared/made accessible for verification and re-use? Outline the data utility: to whom will it be useful. Please contextualize re-use…	Reports for the deliverables. Trend results will be shared in an aggregate and generalised way. Interviewed companies will receive a personal and anonymised positioning report.
Dissemination Level of Data	
Confidentiality/ Sensitive data. If data cannot be made available, explain why. Who will have access?	Data from interviews are confidential and anonymous since they are private data on the manufacturing, production and maintenance characteristics and preferences of companies.
Storage & Disposal	
How will this data be stored?	Deliverable 7.3 and 7.4
How long is it required to keep the data? Expire date. Will revisions be kept? Specify revisions…	At least until the end of the project
Rules of Data Naming	
Standards or internal rules	None
Release Date	
Can data be uploaded in an open research data pilot? When?	No, in case public data is generated, it will be released together with D7.3 and D7.4 (by M42)

5.5.2 Dataset summary linked per WP

In this subsection, a synthesis of the datasets generated in each work package will be shown in order to structure the entire RDI generated on the project.

WP2	Operating system and networked machine simulators	WP leader: INNOVALIA

Objective: In this WP we found mainly datasets that are collected and agglutinated by what we called in Z-BRE4K cognitive embedded systems that are responsible of the data collection, pre-processing and share. Each of them associated to a specific machine or process monitored in different project scenarios. Both specific domain expert knowledge as well as generic pattern recognition dimensionality reduction techniques (PCA, LDA…) will be used to build the Cognitive Embedded Condition Monitoring (CECM) system for real-time feature extraction. Data then will be shared through the IDS connectors that are deployed on the different project scenarios. Through these data sharing channels is how the different DSS, FMECA, Fault Prediction, … components will be gathered data for their exploitation.

Data description: Data generated will be digital and many format data can be accepted (.CSV, .TXT, .JSON, etc). However, data output will be in the main format for Orion Context Broker (.NGSI).

Instrument and tools: Inputs come from the different machines and sensor connected to these cognitive embedded systems.

Terms & Conditions of data generated: Data are collected from the manufacturing processes (end users' collection systems).

WP3	Knowledge and Predictive Modelling	WP leader: BRUNEL

Objective: Data is required for the Z-BRE4K semantic model (ontology) development. The ontology will serve as a common reference model for the annotation and description of the knowledge to represent manufacturing system performance. The ontology will be used in order to drive the semantic framework. Furthermore, it will be used for data integration, visualization, inferencing/reasoning.

Thanks to the understanding of these models, The KBS, FMECA, Failure detection components will be able to understand the machine or process datasets that will read from the OCB.

Data description: In this context, the ontology will be appropriately implemented using standard-based languages as RDF and the generated output will be the semantic enrichment of shop-floor data for representation of processes, actors, alarms, actions, work pieces/products, etc., e.g. RDF Triplets, .CSV, .XML, .JSON.

Regarding the KBS, FMECA, Failure detection components, they will require data on the following formats: .JSON, .CSV, .TXT…

Instrument and tools: In most cases, these components should obtain the data through the Orion Context Broker or directly from the Cognitive embedded system if the solutions are deployed on site.

Data reused: The ontology will be re-used in order to drive the semantic framework. Furthermore, it will be used for data integration, visualization, inferencing/reasoning.

WP4	Design of Strategies and Integration of Intelligence	WP leader: EPFL

Objective: The purpose of the data collection and generation is to facilitate the development of strategies in order to improve maintainability and operating life of production systems, starting with the project use cases. For that aim, the objective is to create Decision Support Systems for predictive and JIT maintenance towards operational optimization. The raw data will be collected from plants, through which it is intended to build, validate and verificate the KPI models to measure the process and machine control optimization and to minimize defect.

Data description: Data will be generated on shopfloor and will be collected mainly digital on the following formats: .CSV, .JSON, .TXT, .PDF and .QIF.

Instrument and tools: Data from the shopfloor will be gathered through MES, SCADA, Production Management Systems and directly from the Cognitive Embedded Systems. Already installed actuators, sensors and financial data from end-users will be used for monitoring and evaluating the KPIs.

Data generated: Data reutilization will be fine-tuned in order for afterwards on-line & real-time application of them and corrective actions will be considered for the production line based on the results received.

WP6	Demonstration activities (TRL 6->7)	WP leader: SACMI

Objective: Once the methodological framework for Z-BRE4K has been implemented and evaluated on WP5, on this WP real-time data will be gathered, processed and analyzed through the operational environments defined on the 3 demonstrators. So, the final objective would be to demonstrate the benefits of the implementation of the Z-BRE4K predictive maintenance models and Z-Strategies on each use case.

Data description: Data will be generated on the shopfloor by sensors and machines. These data will be digital and hard copies and data formats will come as .CSV, .xml, .QIF, .M3 files, … and the outputs as reports that will be generated as .xls, .doc, .ppt, etc.

Instrument and tools: Possibly collected by sensors at a bench top apparatus and optical and physical sensors that can be found describe on WP6 deliverables.

Data re-use: The existing data are re-used as really useful for making quotations, for process study, quality control, traceability and claim answer.

WP7	Valorization and Market Assesment	WP leader: INOVA

Objective: Collecting feedback on the diffusion and interest towards Predictive Maintenance Technologies by interviewing 50 manufacturing companies at the EU level, both SMEs and LE.

Data will come from companies belonging to the following sectors: process industries; consumer goods manufacturing; high-added value parts manufacturing; industrial maintenance company; OEM of low-end machinery; OEM of high-end machinery; energy and transport sector.

The results of this survey will allow to identify the needs and requirements of companies and to compare them with the PdM solutions elaborated by the Z-BRE4K partnership.

Data description: Data will be generated by interviewing Technical, Production and Maintenance directors and responsible people within manufacturing companies. These data will be numerical and qualitative data collected in .xls, .doc, .ppt… format.

Instrument and tools: Possibly collected by Microsoft Forms, Word questionnaire, paper questionnaires.

> **Data re-use:** The existing data are re-used as really useful for better identifying and targeting market segments and maximising exploitation opportunities for partners, in the overall Business Model Canvas that is being generated within WP7. These data, in an anonymised and generalised way, will be contained in D7.3 and D7.4.

In conclusion, data and data management-related challenges under Z-BRE4K are identified and addressed mainly within WP2, WP3, WP4, WP6 & WP7. As described in the proposal, data to be used in the project will include: real-time and historical data related to: (i) the product (desired specifications; quality inspection results, etc.); (ii) the production equipment and environment (e.g. temperature, pressure, vibrations, etc.); (iii) manufacturing/ production and maintenance (e.g. capacity, planning, etc.). Sources of these data will be: existing sensors and actuators (such as sensors embedded in production machinery, quality inspection equipment, etc.), as well as new novel sensors and actuators (such as laser scanning, visual and/or IR cameras, non-contact profilometers, etc.), and enterprise systems. The type of sensors/ actuators and data to be used will be defined and finalized per Z-BRE4K use case on the basis of the required metrics at product and workstation level at single manufacturing stage, and also at multiple stage. Additionally, non-research data collected related to Innovation Management like the IPR registry that includes the IP strategy per Result are confidential and are only stored in the SHAREPOINT and the website private area, with no access rights for members outside the consortium.

5.6 Policies for access, sharing and re-use

Data generated during Z-BRE4K project will be confidential; ownership and management of intellectual property and access will be limited. For this purpose, policies for access, sharing, and re-use have been stablished.

However, considering the description and instructions of the ORD pilot, Z-BRE4K consortium agrees on making publications and the research data related to the scientific publications open. More information about the open data is available in section 4.8.3.1.

5.6.1 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 to 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.

5.6.2 Data Ownership and Access

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.

5.6.3 Naming rules

All data files will be saved using a standardized, 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).

5.6.4 Storage Information

Research data will be stored at the non-free access repository – Z-BRE4K SharePoint - created and maintained by AIMEN. 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 the project, the portal is going to be accommodated with other portals at the same server, so as to minimize the needed costs for its maintenance.

5.6.5 Data sharing and dissemination

Data will be re-used 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.

5.6.5.1 Open Access

Open Access can be defined as the practice of providing on-line 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 (primarily published in academic journals).
- Research data.

All patents and all other publications will require prior agreement in respect to 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 re-use of the findings.

Regarding the research data, only the research data 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. In particular, ZENODO will be used within Z-BRE4K project for this purpose. The data published on ZENODO will then also made available on the OpenAIRE portal that is based on European OA repository infrastructure.

5.6.6 IPR management and security

As an innovation action close to the market, Z-BRE4K project covers high-TRL technologies and aims at developing marketable solutions. The project consortium includes eleven industrial partners from the private sector, in particular, ATLANTIS, INNOVALIA, INOVA, CORE, CRIT, CDS, PHILIPS (demonstrator), SACMI (demonstrator), GESTAMP (demonstrator), TRIMEK and HOLONIX. Those partners obviously have Intellectual Property Rights on their technologies and data, on which their economic sustainability is at stake. Consequently, the Z-BRE4K consortium will protect that data and crosscheck with the concerned partners before every data publication.

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

5.6.7 Data expire date

Copyright statements of the Z-BRE4K 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 digitalized documents, records and other supporting documentation in order to prove the proper implementation of the action and the costs they declare as eligible.

6 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 optimization, which will be collected/generated during Z-BRE4K will be distributed freely.

7   DATA MANAGEMENT PORTAL

The Data Management Portal or internal SharePoint, a web-based portal has been developed within the Z-BRE4K project for the purposes of the management of the research data that will be produced by the project. Regarding the dissemination and exploitation of Z-BRE4K results, the Z-BRE4K website has been created to support these activities: a dedicated section "Public deliverables and dissemination assets" will include the public deliverables, once approved by the project officer, and the scientific papers generated by the project partners within Z-BRE4K context. This website provides up-to-date information about the project. Finally, public data, as aforementioned (deliverables, scientific papers, datasets…), will be published on the open repository ZENODO. Special attention is going to be given on ensuring that the data made publicly available violates neither IPR issues related to the project partners, nor the regulations and good practices around personal data protection.

7.1   Functionalities

On the one hand, the SharePoint is implemented through a web-based platform which enables its users 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, she/he 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 SharePoint will be easily and effectively managed by the members. A variety of graphs, 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 organized 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, Z-BRE4K website (https://www.z-bre4k.eu/) is an open website whose main objective is to disseminate the project purpose, strategies, activities and the outcomes of the research. This communication tool is continuously updated with project progress and dissemination activities/information.

The website has different areas: Project, Partners, News & events, Public Deliverables and dissemination assets, and Contacts. Public deliverables and dissemination assets section will include articles and publications so as to be available for all the visitors.

8 CONCLUSIONS

This report includes the Data Management Plan (DMP) and describes the Research Data Information that is being generated during Z-BRE4K project and the challenges and constraints that need to be taken into account for managing it. In addition, it describes the updated procedures(methodology) and the infrastructure(questionnaires) implemented by the project to efficiently manage the produced data. The DMP is identified as starting point for the discussion with the community about the Z-BRE4K data management strategy and this last version reflects the procedures planned by the work packages at the mid-end of the project.

In this updated version, the elaborated questionnaire has been updated by the consortium partners, including the data they are collecting, and generating or they expect to produce and collect during the project. From this information, it has become clear that currently only WP 2, 3, 4, 6 and 7, are generating or collecting data that can be classified as relevant information according to the definition of the European Commission. As the DMP is not a fixed document, and datasets may differ during the lifespan of the project, the DMP has been again updated and augmented with new datasets and results during this mid-end term of the project, submitting this final version in M32.

Regarding storage information, all the documents generated during the project will be stored in the SharePoint 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 five years after the project completion, as it is described in GA.

On section 4.5.2, a synthesis of the datasets generated in each work package has been performed, in order to structure the entire RDI generated on the project with a brief description of the objectives of the datasets.

9 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.

(http://www.kcl.ac.uk/innovation/research/support/ethics/training/infosheet.aspx)

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. (http://whatis.techtarget.com/definition/metadata)

Or http://www.data-archive.ac.uk/media/54776/ukda062-dps-preservationpolicy.pdf

Repository

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

(http://www.rsp.ac.uk/start/before-youstart/ what-is-a-repository/)

10 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