



HORIZON EUROPE PROGRAMME: HORIZON-CL4-2022-DIGITAL-EMERGING-02

## SoliDAIR

### Solid, rapid and efficient adoption of Data, AI & Robotics applications in production

#### Deliverable D1.3: Data Management Plan

<b>Primary Author(s)</b>	Martin Wifling   ViF
<b>Deliverable Type</b>	Report
<b>Dissemination Level</b>	Public
<b>Due Date (Annex I)</b>	31.3.2024 (Month 6)
<b>Pages</b>	16
<b>Document Version</b>	Final
<b>GA Number</b>	101120276
<b>Project Coordinator</b>	Andreas Frommknecht   Fraunhofer IPA (FHG) (andreas.frommknecht@ipa.fraunhofer.de)

Contributors	
Name	Organisation
Martin Wifling	Virtual Vehicle Research
Tobias Weigand	Brose
Kerman Osoro	CIE
Linghao Zhou	THL
Panagiotis Mavridis	THL
Aksoy Firat	Bosch
Hakan Muslu Cem	Bosch
Christoph Mitteregger	AUTFORCE
Lukas Schreyer	AUTFORCE

Formal Reviewers	
Name	Organisation
Aksoy Firat	Bosch
Hakan Muslu Cem	Bosch
Christoph Mitteregger	AUTFORCE

Version Log			
Rev #	Date	Author	Description
<b>0.9</b>	12.03.2024	Martin Wifling (ViF)	Release Candidate
<b>1.0</b>	22.03.2024	Alexander Katzbeck (AUT)	Quality review
<b>2.0</b>	26.03.2024	Emil Andreas (FHG)	Formatting check
<b>3.0</b>	28.03.2024	Andreas Frommknecht (FHG)	Coordinator review and approval, deliverable ready for submission

## Project Abstract

SoliDAIR aims to accelerate the uptake of Artificial Intelligence (AI) and Robotics in European manufacturing, using Data as an enabler. It will co-develop and demonstrate tailored solutions to digitalise and automate visual inspection and physical testing, enable predictive quality control and process optimisation. The SoliDAIR project tackles the problem of AI & Robotics systems not being extensively used in the production industry, because it is not clear whether they are safe and when or why they will fail, by researching, developing and testing methods that are as solid and trustworthy as possible to be adopted by the European industry, while being cost-efficient to develop and replicate.

New methods and tools will be developed by research and technology providers, which leverage the current state of the art in visual AI, AI for process data, and smart & collaborative Robotics. The developed technologies will be applied and demonstrated in 4 industry use cases to prove their functionality and applicability in real production environments. The objective is to improve production processes through digitalised and automated quality control for high volume, high rate and flexible manufacturing. The developed methods shall be efficiently and easily adaptable and replicable, so they can be easily applied to new use cases outside the consortium.

## Table of Contents

Public Summary .....	4
Introduction.....	5
Rational of this deliverable.....	5
Levels of data .....	6
Dataset descriptions (collected and generated) .....	8
Datasets from the Use-Cases.....	8
BROSE use-case .....	9
CIE use-case.....	10
BOSCH use-case.....	11
AUTFORCE use case .....	12
Conclusions.....	13
References .....	14
Acknowledgements and disclaimer .....	15
Abbreviations and Definitions.....	16

## Public Summary

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be collected and generated by the project.

According to Digital Curation Centre (DCC), a DMP contributes to save time and effort, makes the research process easier, helps to validate if the necessary support is considered, and enables to make sound decisions.

The first issue of the DMP supports project partners to:

- understand the data and use it when needed;
- ensure continuity if project staff leave or new researchers join;
- avoid unnecessary duplication e.g. re-collecting or re-working data;
- contribute to more collaboration and advances research;
- increase visibility and impact;
- manage citations of other researchers on your data.

The present document provides the initial version of the Data Management Plan, a final revision (Deliverable 1.4) will be submitted by M18 of the project. In this initial version, the focus lies on the available input data provided by the use-cases.

The present DMP has been developed based on the EC guidelines on FAIR (findable, accessible, interoperable and reusable) data management in Horizon 2020<sup>1</sup>, the guidelines from the Digital Curation Centre (DCC)<sup>2</sup>, and the Open Access (Open Science) policy lines from the EC.

Due to the nature of the SoliDAIR project and the developed algorithms being based on strictly confidential production data of the partners, the project is not participating in the Open Access (OA) initiative.

---

<sup>1</sup> Guidelines on Data Management in Horizon 2020, DG Research and Innovation, 2016.  
[https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

<sup>2</sup> How to develop a data management and sharing plan, Sarah Jones, DCC, 2011  
<https://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How%20to%20Develop.pdf>

## Introduction

### Rational of this deliverable

This deliverable aims to present the initial version of the Data Management for the SoliDAIR project. The Data Management Plan aims to facilitate the data management of the project, the plan will allow to better understand the data produced and used, to ensure continuity if there are changes in the project's human resources, to avoid duplications i.e. re-collecting or re-working data, to allow validation of results, to contribute to collaboration through data sharing, to increase visibility and greater impact.

This deliverable is produced under the frame of work package 1 Project Management and Coordination. A final update of the data management is expected in the middle of the project in M18.

The sources used to develop the content were the articles defined in the consortium agreement and the contributions from the project partners.

The deliverable is composed of the following sections:

- Objectives and the rationale of the data management plan
- Data set description and allocation (collected and generated) by Use-Cases
- Conclusions

Whereas this current version of the DMP focuses on the understanding of the available data and the handling of such, the second version of the DMP (M18) will be driven by the methodologies developed in the Work Packages and will be in line with the EU Guidelines of FAIR data management:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

## Levels of data

The DMP outlines how the consortium handles data in the project on three levels:

- a) the use case level which will manage the use case specific and need-to-know-based information and data;
- b) the project scope which will handle, store and exchange data on the project level and
- c) public data and project results which will be handled through WP5.

In this version of the DMP, all actual data is assigned to Level 1 data, as the origin of the datasets are (raw) production data with the purpose of understanding the data. Throughout the research process and the methodology development, workshops assigning the specific and generated data to the relevant levels and all subsequent actions are planned (WP2, M12-M15).

**Table 1: Levels of SoliDAIR data**

<b>Type of data</b>	<b>Findability</b>	<b>Accessibility/ Reusability</b>	<b>Interoperability</b>	<b>Curation/ storage</b>
<b>Level 1:</b> Use case data such as sensor and process data, QC data etc.	Data will be held within the use case under use case owners information security / storage policies.	All consortium partners on need-to-know basis. Industrial standards (e.g. Functional Mock-up Units (FMU)) will be used for model exchanges between partners	Ensured through interaction with WP2 and WP5	Depending on WP (e.g., large scale data storage)
<b>Level 2:</b> Data from WP2 such as strategies, frameworks, source code, test data, tools	Managed via in project Sharepoint with suitable access permissions and code repositories like GitHub with links to the project website and partners' repositories	For all consortium partners; available at request to the public, data & code will be shared with 3rd parties through suitable licencing	Will be ensured through quality management measures at project level	FHG, I2M and all partners involved in WP2
<b>Level 3:</b> Documents for publications e.g research papers, guidelines, blueprints	Final analysis of data available through workshops, publications (also using ORE) and project website; source code will be managed via	For all consortium partners; available to the public, data & code will be shared with 3rd parties through suitable licencing like GPL-3.0 or creative common	Ensured through interaction with WP2 and WP5, quality management WP1	All partners involved in WP5

	repositories e.g GitHub	license model CC BY-SA		
--	----------------------------	---------------------------	--	--

Attainment of the objectives and explanation of deviations

There are no deviations from the planning.

## **Dataset descriptions (collected and generated)**

The main purpose of these data sources are to understand the available data, derive the requirements and get the development work started.

This section provides a description of data that are collected and generated during the project for each work package/task at this stage of the project. The description also includes information about the type/format, volume, ownership of the data, etc., as analysed in the following section.

The following tables present the datasets that were collected and/or generated within the relevant work packages until M6.

### **Datasets from the Use-Cases**

All current datasets (M6) are assigned to Level 1 datasets as actual and confidential production data is provided. Hence, the datasets are kept within the Use-Case repositories (see Meta-data descriptions in the Use-Cases).



## BROSE use-case

Topic	Description of Data collected/generated
Dataset name	Brose door modules - initial dataset
Source	Brose Universal End-Of-Line Tester
Scope (WP & Task)	WP3, Task 3.4
Purpose	Limited dataset with images collected from the ring memory storage. Used to setup deep learning model for Brose data and to verify general feasibility. Dataset contains OK samples as well as NOK samples for 2 error classes.
Type/Format	.png files and folder names as labels
Volume	2 GB
Owner	Brose RAS / Brose BAM
Repository	file share via UGS
Language/s	German/English
Confidentiality	Confidential
Link (only if public)	-

Topic	Description of Data collected/generated
Dataset name	Brose door modules - full dataset
Source	Brose Universal End-Of-Line Tester
Scope (WP & Task)	WP3, Task 3.4
Purpose	Dataset from running series production to train and test deep learning model and explainable AI features. To be enhanced with synthetic data if needed. Containing images from 8 End-Of-Line Testers and 6 images per product, with up to 8 error classes. Data is collected between May and September 2024.
Type/Format	.png files and SQL database for metadata
Volume	~ 6-7 TB
Owner	Brose RAS / Brose BAM
Repository	hardware data transfer to FHG
Language/s	German/English
Confidentiality	Confidential
Link (only if public)	-

## CIE use-case

Topic	Description of Data collected/generated
Dataset name	CIE use case private data with THL
Source	CIE and THL individual and joint collaboration contributions
Scope (WP & Task)	Full use-case potential data exchange
Purpose	Historical data, images, defects, classification, equipment options
Type/Format	Any format
Volume	Open
Owner	FCIE + THL
Repository	Fraunhofer's private repository within SoliDAIR's space
Language/s	English
Confidentiality	Confidential
Link (only if public)	-

Topic	Description of Data collected/generated
Dataset name	CIE use case private data with THL
Source	CIE and THL individual and joint collaboration contributions
Scope (WP & Task)	WP3, Task 3.4
Purpose	administrated production dataset for data understanding, training and validation purposes
Type/Format	.csv
Volume	3 GB
Owner	FCIE + THL
Repository	Fraunhofer's private repository within SoliDAIR's space
Language/s	-
Confidentiality	Confidential
Link (only if public)	-

**BOSCH use-case**

Topic	Description of Data collected/generated
Dataset name	Process Parameters Injectors RAW
Source	BOSCH injectors production line
Scope (WP & Task)	WP3, Task 3.4
Purpose	Raw production data to be administrated, cleaned and sorted for timestamps and products
Type/Format	.XLSX
Volume	~ 25-100Mb per dataset
Owner	BOSCH Turkiye
Repository	SoliDAIR sharepoint @ VIF /data/RAW data
Language/s	-
Confidentiality	Confidential
Link (only if public)	-

Topic	Description of Data collected/generated
Dataset name	Process Parameters Injectors RAW
Source	BOSCH injectors production line
Scope (WP & Task)	WP3, Task 3.4
Purpose	administrated production dataset for data understanding, training and validation purposes
Type/Format	.XLSX
Volume	~ 25-100Mb per dataset
Owner	BOSCH Turkiye
Repository	SoliDAIR sharepoint @ VIF /data/training
Language/s	-
Confidentiality	Confidential
Link (only if public)	-

**Note:**

Due to BOSCH customers' requests, a future change of the product type is foreseen. The production of the chosen injector type will be ceased for a design change. Therefore, a switch to another product and datasets seems prudent, as the continuous provision of further datasets of the current product, including purpose made data cannot be assured. The alternative product (injectors) will be chosen asap, however, the strategy of data management (repository, type, etc.) will not be changed.

**AUTFORCE use case**

Topic	Description of Data collected/generated
Dataset name	Process Parameters Gearbox Manufacturing RAW
Source	Gearbox manufacturing line
Scope (WP & Task)	WP3, Task 3.4
Purpose	Historical dataset of production, used for preprocessing
Type/Format	SQL database backup
Volume	190 GB
Owner	External customer
Repository	AUTFORCE/00998 AUTFACTORY/111 SoliDAIR/data/raw
Language/s	-
Confidentiality	Confidential
Link (only if public)	-

Topic	Description of Data collected/generated
Dataset name	Process Parameters Gearbox Manufacturing RAW
Source	Gearbox manufacturing line
Scope (WP & Task)	WP3, Task 3.4
Purpose	administrated production dataset for data understanding, training and validation purposes
Type/Format	.csv
Volume	3 GB
Owner	External customer
Repository	AUTFORCE/00998 AUTFACTORY/111 SoliDAIR/data/cleaned
Language/s	-
Confidentiality	Confidential
Link (only if public)	-

## Conclusions

All Use-Cases have established the relevant data sources derived from actual production components. Initial data samples for data understanding of the research teams have been made available for the teams, have been discussed and data requirements have been analysed for the envisioned methodology developments in WP2.

There are no deviations from the planning.

**Milestone 1 has been achieved and formally approved in the Steering Board (SB) meeting on 8.3.2024 (online).**

## References

<sup>1</sup> Guidelines on Data Management in Horizon 2020, DG Research and Innovation, 2016.  
[https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

<sup>1</sup> How to develop a data management and sharing plan, Sarah Jones, DCC, 2011  
<https://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How%20to%20Develop.pdf>

## Acknowledgements and disclaimer

The author(s) would like to thank the partners in the project for their valuable comments on previous drafts and for performing the review.

#	Partner	Partner full name
1	FHG	FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG EV
2	BRO-B	BROSE FAHRZEUGTEILE SE & CO. KOMMANDITGESELLSCHAFT
3	CIE	FUNDACION CIE I+D+I
4	BOS	BOSCH SANAYI VE TICARET AS
5	AUT	AUTFORCE AUTOMATIONS-GMBH
6	SISW	AUTFORCE AUTOMATIONS-GMBH
7	UGS	UG SYSTEMS GMBH & CO. KG
8	THL	TWI ELLAS ASTIKI MI KERDOSKOPIKI ETAIREIA
9	VIF	VIRTUAL VEHICLE RESEARCH GMBH
10	I2M	I2M UNTERNEHMENSENTWICKLUNG GMBH

### LEGAL DISCLAIMER

Copyright ©, all rights reserved. No part of this report may be used, reproduced and or/disclosed, in any form or by any means without the prior written permission of SoliDAIR and the SoliDAIR Consortium. Persons wishing to use the contents of this study (in whole or in part) for purposes other than their personal use are invited to submit a written request to the project coordinator.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document shall be liable or responsible, in negligence or otherwise, for any loss, damage or expense whatever sustained by any person as a result of the use, in any manner or form, of any knowledge, information or data contained in this document, or due to any inaccuracy, omission or error therein contained.



## Abbreviations and Definitions

Term	Definition
DMP	Data Management Plan
QC	Quality Control
SB	SoliDAIR Steering Board