openZDM

OPEN PLATFORM FOR REALIZING ZERO DEFECTS IN CYBER PHYSICAL MANUFACTURING

Data Management Plan

openZDM

Version	1.0
WP	1
Delivery Date	30 Nov 2022
Dissemination level	PU
Deliverable lead	LMS
Authors	LMS
Reviewers	LMS, INTRA
Abstract	This report describes the initial plan for the management of the data expected to be acquired or generated during the openZDM project along with the approach that will be followed to preserve them in a structure way facilitating their maintenance, update as well as sharing, following where possible the principles of FAIR data.
Keywords	Data, FAIR, Initial, Management, Plan
License	cc BY ND This work is licensed under a Creative Commons Attribution-No Derivatives 4.0 International License (CC BY-ND 4.0). See: https://creativecommons.org/licenses/by-nd/4.0/

Dissemination Level:	
PU	Public, fully open
SEN	Sensitive, limited under the conditions of the Grant Agreement
Classified R-UE/EU-R	EU RESTRICTED under the Commission Decision No2015/444
Classified C-UE/EU-C	EU CONFIDENTIAL under the Commission Decision No2015/444
Classified S-UE/EU-S	EU SECRET under the Commission Decision No2015/444
Type	
R	Document, report (excluding the periodic and final reports)
DEM	Demonstrator, pilot, prototype, plan designs
DEC	Websites, patents filing, press & media actions, videos, etc.
DATA	Data sets, microdata, etc.
DMP	Data management plan
ETHICS	Deliverables related to ethics issues.
SECURITY	Deliverables related to security issues
OTHER	Software, technical diagram, algorithms, models, etc.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUROPEAN HEALTH AND DIGITAL EXECUTIVE AGENCY (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Version History

Version	Date	Owner	Author(s)	Changes to the previous version
0.1	2022-09-25	LMS	LMS	Outline
0.8	2022-11-19	LMS	LMS	Full draft
0.9	2022-11-29	LMS	INTRA	Reviewed draft
1.0	2022-11-30	LMS	LMS	Final draft/Submitted version

Table of Contents

Version History 3

Table of Contents 4

List of Abbreviations & Acronyms 5

List of Figures 6

List of Tables 6

Executive Summary 7

1 Data summary 8

2 FAIR data 9

2.1 Making data findable, including provisions for metadata 10

2.2 Making data accessible 10

2.3 Making data interoperable 11

2.4 Increase data re-use 11

3 Other research outputs 12

4 Allocation of resources 12

5 Data security 13

6 Ethics 13

7 Other Issues 13

8 Conclusions 13

References 14

List of Abbreviations & Acronyms

AI	:	Artificial Intelligence
API	:	Application Programming Interface
ASCII	:	American Standard Code for Information Interchange
CAD	:	Computer-Aided Design
CERN	:	Conseil Européen pour la Recherche Nucléaire (European Organization for Nuclear Research)
DMP	:	Data Management Plan
DOI	:	Digital Object Identifier
EC	:	European Commission
FAIR	:	Findable Accessible Interoperable Reusable
ICT	:	Information and Communication Technology
JSON	:	JavaScript Object Notation
ML	:	Machine Learning
NDI	:	Non-Destructive Inspection
OAI-PMH	:	Open Archives Initiative Protocol for Metadata Harvesting
R&D	:	Research & Development
REST	:	REpresentational State Transfer
URL	:	Uniform Resource Locator
UTF	:	Unicode Transformation Format
WP	:	Work Package

List of Figures

Figure 1: openZDM community in ZENODO 9
Figure 2: openZDM community identifier on ZENODO repository 10

List of Tables

Table 1: Table of indicative data formats per data file 8
Table 2: Information accompanying each upload to make the data findable. 10
Table 3: Indicative readme file elements 11

Executive Summary

The purpose of this document is to provide the initial plan for managing the data acquired, collected, and/or generated during the lifetime of the openZDM project and its activities. As such, and following the template provided, the openZDM DMP includes the following aspects:

- Types of data considered in the context of the project
- Supported data formats, metadata, and management approach for proprietary formats
- Access and sharing principles and policies
- Policies for making data findable, accessible, interoperable and re-usable
- Data storage and preservation
- Information accompanying the uploaded data, such as compilation guidelines and context
- Responsibilities and cost
- Associated risks concerning ethical and legal aspects as well as management of personal data and software

The openZDM project will use the ZENODO repository to upload and share data with the wider public via its working space uniquely identified as “openzdm”. The ZENODO repository is hosted in the servers of CERN and is expected to be operational for approximately the next 15(fifteen) years, and to the best of the consortium's knowledge, complies with the FAIR principles following widely accepted by the research community (meta)data standards.

The openZDM consortium considers making data publicly available based on its activities in the project and after making sure there is no confidentiality concern or risk with making the corresponding data publicly available. This will be validated with all consortium members before uploading any data.

In addition, the process of curating and maintaining the data as well as the ZENODO community will be the responsibility of the coordinator. Finally, each upload associated with a unique DOI will be linked to the EC Participants Portal.

1 Data summary

In the context of the **openZDM** project, new data will be created to support its R&D activities, such as measurements, data models, and point clouds. Furthermore, existing data will be reused, such as historical data from the end-users and maintenance records to facilitate the training of machine learning models for data analysis and prediction.

At the moment of compiling this report, all data that are relevant to the R&D activities of the project and coming from the products/processes that have been identified as in the scope of the project are considered and not excluded. It is estimated that in the process of analysing them, some will be less relevant for the development and validation activities, and thus will be rejected from the procedure.

Thus, the data types that will be created and/or worked within the context of the project are presented below:

- **Observational**, such as human feedback on the results of the project solutions.
- **Experimental**, such as the data collected through the NDI systems of the project
- **Compiled** data from the project software, such as data analytics.
- **Simulation**, derived by the project digital twin and connected simulation tool.

Regarding the supported data formats and considering that research data of the above types may come in various formats, the project consortium aims to focus on non-proprietary and open formats with standard character encodings (such as ASCII and UTF-8). In case of non-compliance, then conversion to corresponding open formats will be considered and applied wherever possible.

A preliminary set of data formats expected to be used in the project activities per type of data file is presented in the below table (Table 1).

Table 1: Table of indicative data formats per data file

Data file/content	Format
Text	TXT, XML, LATEX, TEX, RTF, DOC, DOCX, PDF
Graphics	TIFF, JPG, GIF, PNG, BMP | SVG, SWF, EPS, CGM
Papers/ Books/ Digitized instructions/ etc.	PDF, HTML
Digital audio	MP3, MP4, WAV, WMA, M4A, FLAC
Digital video	MP3, MP4, MOV, AVI, WMV, OGV
Spreadsheet	XLS, XLSX, ODS, CSV, TSV, TXT
Database	FIC, XLS, XLSX, IBD, MAR, DDL, ACCDC, SQLITEDB, MDB, MDF, DB, ODB, SQL, FDB, DBS, XLD, FRM, MPD,
Presentation	PDF, PPT
Compressed	ZIP, RAR, ISO, 7Z, TAR
Statistical data	TXT, CSV, JSON, SAV, SPS, XLS, XLSX, TAB, DBF, DWG, SHP, SHX, MDB, RTF, XML, TIF, DTA, R, RMD, DBASE
CAD	DWF, DXF, DWG, DWFX, DWT, SLDPRT, PSS, IGS, IFC, IGES, PLT, STL, CATPART, DC3
Point clouds	XYZ, PLY, OBJ, BTS
Software	obfuscated/compiled/executable

The openZDM project targets the greater objective of reducing manufacturing defects and wastes, through the use of advanced digital solutions including the use of AI. As such, large volumes of data (estimated from some GBs to TBs) will need to be collected to support the development of the required AI methods and models, for analytics, classification, and/or prediction, as well as the envisioned digital twins. Furthermore, considering that the process of data processing, cleaning, and analysis is a time-consuming procedure, it is expected that at the beginning more data will be collected than the ones necessary for the objectives of the project, such as more parameters measured for the training of an ML model than the minimum required set.

In addition, synthetic data are expected to be generated to complement the real-world datasets thus creating rich and representative datasets of the observed situation facilitating the use of ML techniques.

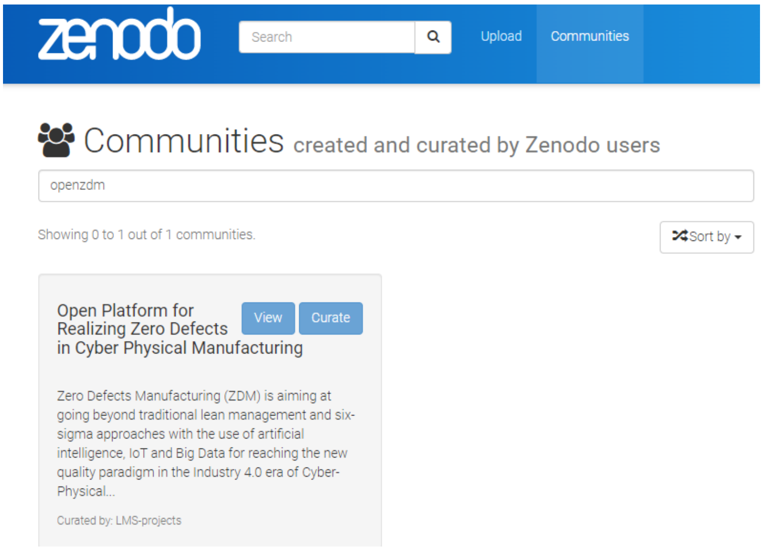
The openZDM datasets are expected to be useful to relevant sectors and researchers to support their activities. Considering that the openZDM project covers a wide spectrum of industrial sites and processes, with a wide range of digital solutions, the data required to enable those solutions can serve as a basis for further activities.

Nevertheless, the data, directly acquired from manufacturing sites, are expected to withhold confidential aspects of the corresponding production system. Hence, any potential dataset identified will be checked for any possible confidentiality issues and if needed anonymised before publication or restricted to the consortium only.

Last but not least, the project open access outputs will be reported in the EC Participant Portal.

2 FAIR data

In compliance with the principles of making data reusable for the scientific community, first published in 2016 [1], the openZDM project will use the ZENODO repository for uploading and sharing data as well as material that is characterised by the consortium as publishable to the wider public audience. Towards this, a specific community has been created for uploading data and material that will comply with the aforementioned restriction, meaning to have been previously approved by the consortium members as publishable to the wider public. The community has been assigned a unique identifier, namely "*openzdm*".



Communities created and curated by Zenodo users

openzdm

Showing 0 to 1 out of 1 communities.

Open Platform for Realizing Zero Defects in Cyber Physical Manufacturing

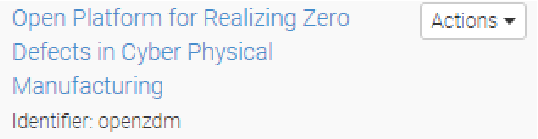
Zero Defects Manufacturing (ZDM) is aiming at going beyond traditional lean management and six-sigma approaches with the use of artificial intelligence, IoT and Big Data for reaching the new quality paradigm in the Industry 4.0 era of Cyber-Physical...

Curated by: LMS-projects

Figure 1: openZDM community in ZENODO

Furthermore, the openZDM community has been assigned a unique identifier, *"openzdm"*, accessible via the following URL:

https://zenodo.org/communities/?p=openzdm


Open Platform for Realizing Zero Defects in Cyber Physical Manufacturing

Identifier: openzdm

Figure 2: openZDM community identifier on ZENODO repository

Lowercase letters have been selected for the identifier of the community aiming to make it more findable by anyone interested in the future.

The compliance of the ZENODO repository, thus the data shared through its communities, with the FAIR principles can be reviewed in greater detail in [2].

2.1 Making data findable, including provisions for metadata

Data uploaded to the "*openzdm*" community in ZENODO will be identified by a persistent identifier followed by rich metadata. ZENODO assigns a unique DOI to all publicly available uploads (datasets, papers, videos, or other) in a specific community to make both the upload and sharing easy and uniquely citeable. Furthermore, the identifiers of the community, or else the shared workspace, and uploads remain as long as the community exists.

In particular, each upload to the ZENODO community can be accompanied by the information provided in the following table (Table 2). In addition, the metadata is compliant with DataCite's Metadata Schema [3].

Table 2: Information accompanying each upload to make the data findable.

Required	**Recommended**	**Optional**
Type of upload	Funding	Contributors
Basic information: • DOI • Publication Date • Title • Authors • Description • Version • Language • Keywords • Additional notes	Communities in which the upload will appear	References
		Journal
		Conference
		Book/Report/Chapter
License		Thesis
Access rights	Related/Alternate identifiers	Subjects

The openZDM consortium will provide as a minimum the required and recommended information and the optional will be case-dependent.

2.2 Making data accessible

Apart from their unique identifier, data are retrievable using the OAI-PMH protocol as well as through the public REST API. Authentication and authorisation are supported by the protocol itself for metadata shared on the public web.

Data and metadata will be accessible for the entire lifetime of the repository, hosted by CERN, and expected to last at least 15(fifteen) more years. In addition, the servers where data are hosted are separated from the ones hosting the metadata.

Regarding the openZDM data, in principle, only available uploads will be considered, without any fee or embargo requirement. Thus, the openZDM consortium will examine and verify the anonymisation or the lack of any confidentiality concern from any data uploaded to the repository and linked to the openZDM Grant Agreement.

Furthermore, demo software/models are made publicly available, but this is to be decided in the course of the project and based on the available time of the consortium members and the confidentiality concerns related to such actions. Nevertheless, if supported, the software will be obfuscated, with anonymised data, in an executable format, pre-configured to demonstrate certain features of the actual tool.

No open-source code is expected to be publicly or otherwise shared outside of the consortium members, during or after the project's lifetime.

2.3 Making data interoperable

Interoperability will be followed by the JSON schema, which is applied in the ZENODO repository, as an internal format. Additional formats are also supported for data export such as Dublin Core.

Furthermore, metadata can be linked to external metadata by a URL and may use vocabulary compliant with the FAIR principles.

2.4 Increase data re-use

The data uploaded in the "*openzdm*" community, complying with the ZENODO requirements, are expected to make the data reusable by any domain-relevant stakeholder, through the required, recommended, and optional entries according to the mandatory requirements of DataCite.

In addition, data uploaded to the openZDM community in ZENODO will be accompanied by a README.txt file that will include basic information, describing the data, enriching the ones provided by ZENODO. The basic entries of a readme template are presented in Table 3.

Table 3: Indicative readme file elements

Element Name	**Purpose**
subject	The root name of each upload starting with openZDM_
Description	A textual description of the data and their purpose/objective
origin	It defines the data source
volume	The uploaded volume for the validity check
contents	A list of the files uploaded and their format
Additional information	This field provides any additional instructions related to the data, for example, the compilation process for an experiment with the data provided and to repeat the findings of a publication. In addition, information about their re-use will be included such as the purpose, functionality or.
Limitations	Limitations and/or warnings stemming from the uploaded files to a future user and concerning the objective the data were created for.

Thus, the addition of the README file is expected to make the data reusable by any interested party in the future and after the end of the project, based on the standards provided by the ZENODO repository.

All data collected or generated in the project to facilitate its activities are expected to follow the four steps of the quality assurance cycle, Plan, Do, Check, and finally Act, designed by their creator(s) and validate the responsible parties of the consortium in charge of using the data. Hence, the data identified as candidates for open access will undergo a systematic approach to data cleansing and assess their quality and validity before publishing them in the "openzdm" community. In addition, even though the openZDM project does not collect any personal data, the ethical aspect of all data to be made publicly available will be investigated in collaboration with consortium partners as well as any potential security risk.

However, upon publication of the data, it is expected that no confidential or harmful for the consortium or any other party has been identified. However, the consortium has no responsibility for any negative impact from the use of the data made publicly accessible in the ZENODO repository. The use of the data will be the responsibility of the party using them.

Apart from the ZENODO portal, all data remain stored in the consortium-shared repository as required by the Grant Agreement. Furthermore, all publicly accessible data available to the "openzdm" community will be maintained on the servers of the coordinator, after the end of the project re-uploading any missing/corrupted uploads that come to its attention and with the consensus of the consortium or the specific partners involved in the corresponding data creation.

The management of the open access data, their cleansing, preservation as well as the communication with the partners will be under the responsibility of the project coordinator, among other activities in the project.

3 Other research outputs

In addition to data, it is expected to make publicly accessible the following research outputs:

- Open access publications
- Videos that have been approved by the consortium for public access
- Demo executable software/models, preconfigured and as executable
- Demo simulation models/digital twins, preconfigured and mostly for training purposes

In all the above material, anonymisation and/or obfuscation will be carried out in advance of uploading them to the ZENODO community along with the consensus of the consortium. The above-mentioned principles followed for data uploaded to the repository will apply to other research outputs as well.

4 Allocation of resources

The resources required for the data curation, cleaning, anonymisation, and consortium consensus acquisition will be allocated mainly by the coordinator, keeping the contribution of the consortium partners to a minimum, mostly for the confidentiality approval procedure before publishing. For the latter questionnaires may be used to collect feedback related to ethics, legal aspects, and personal data. Instead of questionnaires, Microsoft forms may be used to speed up the entire procedure while maintaining proof of the whole procedure.

The data acquisition and collection procedure is expected to be part of the R&D activities of the project, thus no extra effort is expected.

5 Data security

No additional provisions that the ones already in place by the ZENODO repository will be considered by the consortium. However, upon making the data publicly accessible any risk is expected to have been eliminated.

Regarding data recovery and as an additional precaution, the coordinator will also store the publicly available data on its servers. In case of need or request, and with the consensus of the consortium or the partners related to the specific uploads, the data may be re-uploaded to the ZENODO repository.

6 Ethics

No legal or ethical aspects are expected to apply upon uploading the data to the ZENODO repository. In particular, the consortium will assess the ethical and legal risks that might occur from uploading data or other material, and if probable, the upload will be discontinued.

Hence, only material free of such concerns will be uploaded, providing open access and free to re-use, with the required instruction and guidance for any interested party.

However, the consortium takes no responsibility for the use of the uploaded data from any other party during the project or after it. The use of the material will be under the sole responsibility of the people/organizations downloading them from the ZENODO repository and using them.

A similar notification will accompany every upload to the ZENODO repository, clarifying any potential legal issue that anyone might want to raise in the future.

7 Other Issues

No other relevant procedures that should be taken into consideration have been identified or foreseen. In case of any relevant procedure remerges or comes to the attention of the consortium this will be included in the updated version of the DMP.

8 Conclusions

This deliverable provides the initial DMP of the openZDM project concerning the (meta)data expected to be collected or generated in the project's lifetimes. In addition, the repository facilitating storage and sharing has been identified, in compliance with the FAIR principles. Furthermore, the approach to be followed for making data publicly available has been described. Any change or modification to what is included in this deliverable will be included in a future updated DMP.

References

[1] M. D. Wilkinson, “The FAIR Guiding Principles for scientific data management and stewardship.” *Scientific Data*, vol. 3, no. 1, 2016, doi: 10.1038/sdata.2016.18.

[2] https://about.zenodo.org/principles/

[3] https://schema.datacite.org/

openZDM