

Project acronym: Lasers4MaaS

Project title: Laser-as-a-Service Digital Platform with Dynamic Beam Shaping for Acceleration of Smart, Decentralised and Sustainable Factory of the Future

Call HORIZON-CL4-2024-TWIN-TRANSITION-01-03 Manufacturing-as-Service: technologies for customised, flexible, and decentralised production on demand

Grant agreement No: 101178719

Work-package 20: Management and coordination (reporting period #1)

Deliverable D20.1: Data management plan

Owner: WMG

Due date: 28th February 2025

| 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 | x |
| ETHICS | Deliverables related to ethics issues | |
| SECURITY | Deliverable related to security issues | |
| OTHER | Software, technical diagram, algorithms, models, etc | |

| Dissemination level | | |
|----------------------|--|---|
| PU | Public, fully open, e.g. project website | x |
| SEN | Sensitive, limited under the conditions of the Grant Agreement | |
| Classified R-UE/EU-R | EU RESTRICTED under the Commission Decision No2015/444 | |

Disclaimers: this project has received funding from the European Union's HORIZON EUROPE research and innovation programme under grant agreement No. 101178719. The content of this publication is the sole responsibility of the Consortium partners listed herein and does not necessarily represent the view of the European Commission or its services.

HISTORY OF CHANGES

| Version | Publication date | Change |
|----------------|-------------------------|--|
| 1.1 | 02/02/2025 | First draft |
| 2.0 | 15/02/2025 | Second draft |
| 3.0 | 24/02/2025 | Final version approved by the consortium |

LIST OF AUTHORS

| Name | Organization | Email |
|--------------------|---------------------|--|
| Pasquale Franciosa | WMG | p.franciosa@warwick.ac.uk |
| Volkher Onuseit | USTUTT | volkher.onuseit@ifsw.uni-stuttgart.de |
| Marco Cinelli | ULEI | m.cinelli@luc.leidenuniv.nl |
| Franco Donati | ULEI | f.donati@cml.leidenuniv.nl |
| Constantin Zenz | TUW | constantin.zenz@tuwien.ac.at |
| Antti Partanen | FUTURICE | antti.partanen@futurice.com |

REVIEWED BY

| Name | Organization | Email |
|--------------------|---------------------|--|
| Pasquale Franciosa | WMG | p.franciosa@warwick.ac.uk |

TABLE OF CONTENT

1. Executive summary5

2. Introduction and background6

3. Data summary6

3.1. Data type and format6

3.3. Expected size of the datasets7

3.4. Data value7

4. FAIR principles8

4.1. Making data findable, including provisions for metadata.....8

4.1.1. *Persistent identifiers*.....8

4.1.2. *File naming conventions*.....8

4.1.3. *Keywords and metadata*8

4.2. Accessibility.....9

4.2.1. *Availability and accessibility of data*.....9

4.2.2. *Personal and sensitive data*.....9

4.2.3. *Expiry data for open data*.....9

4.3. Interoperability.....9

4.4. Reusability10

4.5. Allocation of resources10

4.6. Data security11

5. Ethics11

LIST OF TABLES

Table 1: Anticipated project data types and formats in Lasers4MaaS 6
Table 2: Example of file naming that will be adopted in Lasers4MaaS 8
Table 3: Data availability and re-use in Lasers4MaaS..... 10

LIST OF ABBREVIATIONS

| | |
|-----------------|--|
| AI | Artificial Intelligence |
| CA | Collaboration Agreement |
| DMP | Data Management Plan |
| DOI | Digital Object Identifier |
| DPP | Digital Product Passport |
| GA | Grant Agreement |
| GB | Gigabytes |
| IP | Intellectual Property |
| KEMs | Key Enabling Methodologies |
| MDS | Manufacturing Data Spaces |
| ML | Machine Learning |
| OPC | Open Platform Communications |
| RAMI 4.0 | Reference Architecture Model Industrie 4.0 |
| TB | Terabytes |
| TRL | Technology Readiness Level |

1. Executive summary

This deliverable constitutes the initial Open Science and Data Management Plan (DMP) for the Lasers4MaaS project. The Lasers4MaaS's DMP follows the provided EU DMP template. Data Governance and handling of data in the project is stipulated in the form of initial project guidelines, the various data types expected to be generated in the project, and whether and how it will be made open and accessible for verification and re-use (FAIR principles). This DMP will be revised throughout the entire project duration and updated, if necessary, by the project coordinator (WMG). This version of the DMP has disclosed the plans for data management according to best available information at project month M3. While beneficiaries are encouraged to look for ways to share data with the research community, they are under no obligation to disclose data if this goes against their interests and commercial confidentiality.

Openly available data: data are subject to a case-per-case evaluation to determine whether they should be made openly available to protect industrially/commercially sensitive information. Only selected data will be shared after internal consultation and approval in the consortium. The general rule will be “data must be “as open as possible, as close as necessary.” If data is openly shared, Lasers4MaaS will use the Zenodo platform - <https://zenodo.org/>. Each partner must ensure open access to all peer-reviewed scientific publications that proceed from their research outcomes. This can be through local institutional repositories. Public deliverables as well as open access publications will be uploaded on the Lasers4MaaS website.

Internal data: General rules for internal data sharing follow the CA and GA signed by all project partners. IP arrangements will be discussed in due course and, if necessary, data be licensed using standard reuse licenses, in line with the obligations set out in the CA and GA. Lasers4MaaS is currently using Microsoft TEAMS as an intranet repository for project-related data and management. It has the advantages of chats, meetings and seamless sharing of documents, amongst the various features. The project folder on TEAMS applies a strict policy in granting and revoking access to data.

2. Introduction and background

This document is the Lasers4MaaS’s DMP which will be revised throughout the entire project duration. The present DMP is an indicative plan as to what kind of data the project beneficiaries expect to generate during the project, and how these data will be managed. This DMP follows the EU standard DMP template. The objective of Lasers4MaaS is to revolutionise laser welding by dynamic beam shaping and digital technologies for servitisation of manufacturing. Building upon the growing advancement in dynamic laser beam shaping as “all-in-one” laser welding tool, with immense potential for servitisation, Lasers4MaaS introduces a six-point strategy to reconfigure, connect, control, predict, improve and ensure compliance in manufacturing. With demonstrations in sectors like automotive, aerospace, food packaging and renewable energy this Lasers4MaaS aligns closely with Green Deal goals.

Table 1: Anticipated project data types and formats in Lasers4MaaS

| No. | Data Type | Data Format |
|-----|--|---|
| 1 | Technical specifications, KPIs | *.csv, *.xlsx, .docx, *.pdf |
| 2 | Sensor data (CMOS/CCD camera, infrared LW camera, OCT, photodiodes, microphones, etc.) | *.csv, *.xlsx |
| 3 | Microstructures (images, etc.), | *.bmp, *.jpg, *.png |
| 4 | Time series of process data (process inputs and emissions, evolutions of the keyhole dynamics in time, etc.) | *.csv, *.xlsx |
| 5 | High speed camera data | *.mraw, *.avi, *.mpeg4, *.mp4, *.wmv |
| 6 | Simulation data (i.e., 2D and 3D distributions of temperature, etc.) | *.csv, *.txt, *.hdf5 |
| 7 | ML/AI models | *.JSON, *.xml |
| 8 | LCA and cost data | *.csv, *.txt, *.xlsx *.json *.xml |
| 9 | Preferences of users of the Decision Support System | *.csv, *.txt, *.xlsx |
| 10 | Code files of digital platform | *.ts, *.js, *.tsx, *.jsx, *.py, *.toml, *.yaml, *.html, *.css, *.java, *.sql |
| 11 | CAD data, system specifications, measurement results | *.stl, *.stp, *.igs, *.parasolid |
| 12 | Website traffic; audience engagement statistics from social media | Exports from LinkedIn analytics (*.xlsx and/or *.pdf), photos and/or videos of project participants (*.bmp, *.jpg, *.png and *.mp4) |
| 13 | Project management data (e.g. financial, resources) | Financial tables (*.xlsx), reports and deliverables (*.docx, *.pdf) |

3. Data summary

3.1. Data type and format

During the project, data will be generated for designing, testing and validating the Lasers4MaaS’s platform. The purpose of the data collection/generation is subdivided into the following project phases: phase#1 - development of the key enabling methodologies (KEMs) directly linked to the 6 strategic objectives; phase#2 - validation in lab of the KEMs in lab settings at TRL 4; phase#3 - hardware/software integration and technology validation in relevant industrial environment to achieve TRL 5; phase#4 - technology demonstration in relevant industrial environment to reach TRL 6. It is expected that large data sets with

different types and formats will be generated. Although we may not be able to fully populate the list of data types and formats at this stage, general remarks can however be made. Generally, the consortium will endeavour to utilise common data formats. In some cases, particular laboratory equipment or software may output data in proprietary formats, but where sharing of these data with other partners is necessary, a conversion to or a summary in common formats will be made possible. Table 1 shows the anticipated data types and formats.

The generated data will be quantitative, qualitative and semi-quantitative. They will be obtained by lab-based experiments, computer-based simulations, as well as interviews and surveys. They include sensors data (CMOS/CCD camera, infrared LW camera, OCT, photodiodes, microphones as well as tactile sensors, etc.), microstructures (images, etc.), time series of process data (process emissions, evolutions of the keyhole dynamics in time, etc.), and simulation data (2D and 3D distributions of quantities such as temperature, etc.). Lab-based data will comprise different formats: *.xlsx, *.csv, *.txt, *.avi, *.png/jpg/tiff, *.dicom. Computer-based simulation data will include both outputs from commercial simulation suites (e.g., Comsol Multiphysics, Flow3D, etc.) and open-source packages (e.g., OpenFOAM, etc.). In order to generate reusable and interoperable data, simulation data will be exported in tabular formats (*.csv, *.txt), or open standard file formats such as JSON.

Depending on the amount of generated data from experiments and simulations methods like temporal and spatial coarsening, post-processing, compression or curation can be applied prior to making data publicly accessible, to avoid unsensible use of storage capacities. The ML/AI models will be made available (weight, model architecture, etc.) via open formats (i.e., Open Neural Network Exchange (ONNX)). The LCA data will be available in open and standard formats (*.csv, *.txt, *.xlsx, etc.). The preference of users of the DSS will be collected in different formats (*.csv, *.txt, *.xlsx) and then harmonized according to the project objectives.

The back-end data will be stored on local servers located at partners beneficiaries. To prevent data loss, data will be stored on dedicated PC/work-stations with failsafe RAID systems.

3.2. Reuse of existing data and origin of data

Most of the data in the project will be generated in-project. The project might use existing data (public repositories) for research to broaden and expand the existing databases from the delivery of the project. This is especially applicable to train ML/AI models. The re-used data will come from external and open sources databases as well as from internal databases from the project partners. Databases provided will be assessed beforehand for their suitability by the respective partner. Only technically and legally suitable data from external databases is used in the project. The use of data from external datasets will be acknowledged in any publications produced by the project partners. For the development of the DSS, the existing repositories that will be used include the MCDA-MSS (<http://mcdamss.com>), as well as mcda.it (<http://mcda.it>).

3.3. Expected size of the datasets

At this state of the project, the expected size of the data is unknown, however, general guidelines are provided. For example, high speed imaging videos or X-ray recordings from in-situ laser experiments may reach several Terabytes (TB). Raw data generated in transient high-fidelity simulations can reach 100 GB up to 1TB for a single simulation. With a high number of simulations and experiments expected to be performed, careful consideration will be given on how the data will be prepared for long term storage (temporal and spatial resolution, post processing of raw data, curation of data, etc.) to avoid wasteful use of data storage.

3.4. Data value

The data generated in Lasers4MaaS will be useful for the project beneficiaries, for other research projects in a similar field and for companies that want to develop products or systems of a similar nature.

4. FAIR principles

The research data generated by Lasers4MaaS shall be “FAIR”: findable, accessible, interoperable, and re-usable. It is acknowledged that many data generated in Lasers4MaaS are subject to intellectual property rights or are part of confidential deliverables and therefore cannot be published. The general rule applies: “as open as possible, as closed as necessary”. Data published in public deliverables are made freely available if there is consent of all project partners. The research data will be shared on a cloud-based repository with a general-purpose open-access policy (Zenodo platform - <https://zenodo.org/>). During the project, the “Welding Process Analytics” will be populated with relevant data to aid developing the Lasers4MaaS platform - more details will be available in D5.2: architecture of the Welding Process Analytics Database. The “Welding Process Analytics” will be only accessible to the project partners during the project duration. IP and commercialization routes will be reviewed in the exploitation plans (D17.2 and D19.3).

4.1. Making data findable, including provisions for metadata

4.1.1. Persistent identifiers

Unless any project beneficiary applies persistent identifiers for their own use, no persistent data identifiers are envisaged for the purposes of information sharing between partners in the project. The project manager will ensure that the data is shared among the partners in a safe and organised way using the management tools defined in D20.2: project and risk management plans. Any datasets made available as open data will use the Digital Object Identifier (DOI), for example provided by the cloud-based repository (i.e., Zenodo platform).

4.1.2. File naming conventions

To ensure seamlessly data management, the consortium will adopt a standard for naming internally shared documents as follows: *Index_DocumentName_Partner_Date_Version.ext*.

Table 2: Example of file naming that will be adopted in Lasers4MaaS

| Description | Meaning | Example |
|--------------|---|------------|
| Index | Index number | WP10 |
| DocumentName | Brief name suitable to identify document | TestData |
| Partner | Identifier name for the partner generating the document | WMG |
| Date | Date of document creation | 25/10/2025 |
| Version | Document version number | v1.0 |
| Ext | File extension | txt |

4.1.3. Keywords and metadata

Depending on the underlying data, “search keywords” will be assigned. This will enhance findability and reuse of the datasets. Metadata will be also made available when necessary. Lasers4MaaS will use a structure metadata for those datasets that are published in public catalogues and data repositories. The following metadata will be considered:

- **Scope:** for what purpose was data created/collected?
- **Date of dataset generation:** when was the data generated?
- **Material and methods:** materials used and welding/process conditions (laser type, optical properties, cooling shielding, etc.)
- **Parameter settings:** which settings (both experiment and simulation) were set while generating the data?
- **Name and version of the software used:** which software and version were used?
- **Data type:** does the dataset contain raw data or processed data or both?

- **Variable names:** variable names/parameters are explained or self-explanatory (i.e., defined in the vocabulary of the research field).
- **Data version:** the version of the archived and/or re-used data is clearly specified and documented. Any code (Python, MATLAB, etc.) necessary to manage and retrieve research data will be made available on open-source repositories (GitHub - <https://github.com/>) in order to facilitate data sharing and long-term validity. Once data will be generated, the list of metadata will possibly be adapted in further versions of the Data Management Plan.

4.2. Accessibility

4.2.1. Availability and accessibility of data

Openly available data: data are subject to a case-per-case evaluation to determine whether they should be made openly available to protect industrially/commercially sensitive information. Only selected data will be shared after internal consultation and approval in the consortium. The general rule will be “data must be as open as possible, as close as necessary.” If data is stored at the openly accessible repository (Zenodo), then this repository will have the possibility to enable users to reach selective quality-controlled datasets and versions approved by data owners. Each partner must ensure open access to all peer-reviewed scientific publications that proceed from their research outcomes. This can be through local institutional repository (i.e., WRAP for Warwick).

Internal data: General rules for internal data sharing follow the CA and GA signed by all project partners. IP arrangements will be discussed in due course and, if necessary, data be licensed using standard reuse licenses, in line with the obligations set out in the CA and GA. Lasers4MaaS is currently using Microsoft TEAMS as an intranet repository for project-related data and management. It has the advantages of chats, meetings and seamless sharing of documents, amongst the various features. The project folder on TEAMS applies a strict policy in granting and revoking access to data. Public deliverables as well as open access publications will be uploaded on the Lasers4MaaS website.

4.2.2. Personal and sensitive data

No personal or sensitive research data is planned to be generated or shared in Lasers4MaaS. For non-research personal data, project partners are requested to fill out a consent form regarding the use of photo or video material featuring still or animated images of individuals – in conformity with data protection regulation.

4.2.3. Expiry data for open data

Once approved for publication as open data, the consortium does not envisage an expiry date.

4.3. Interoperability

Interoperability is the heart of Lasers4MaaS and will be continuously nurtured during the project and data interoperability principles will be applied to research data too. Data (both lab-based experiments and computer-based simulations) will be collected and generated following very rigorous and documented protocols, which will be reviewed on a regular basis (at least quarterly) with the project team. For example, lab-based experiments will be recorded with detailed information/metadata about scope of the experiment, time/date, testing procedure, parameters considered during the experiments, etc. All metadata will be stored alongside the generated dataset. Additional information will include documentation and “readme” file with details about how to retrieve and visualise the data, and related cross-link/references to external libraries and datasets. For code and software, a user manual with detailed references to functions and sub-routines will be developed and made available.

Lasers4MaaS, through its objectives, will build on relevant existing standards for manufacturing systems, aligned with the RAMI 4.0 reference architecture that interlink standards with appropriate model layers

then contributing to data interoperability via harmonised protocols and distributed/centralised ledgers. Lasers4MaaS will help shaping new standards and regulations towards upcoming EU initiatives like the digital product passport (DPP) and manufacturing data spaces (MDS). Lasers4MaaS will manage interoperability by: (a) adopting a hybrid approach with distributed and centralised ledgers for data interoperability across factory boundaries - sensitive data will be stored on private local servers (edge devices) while the unclassified parts are sent to the cloud server; (b) developing unified and harmonised protocols and following reference architectures for the manufacturing industry such as Reference Architecture Model Industrie 4.0 (RAMI 4.0) and Open Platform Communications (OPC). More details about the interoperability standards adopted in Lasers4MaaS are in D1.2 (current standards and approaches for interoperability of digital twins for automation workflows) and D.1.3 (current standards and approaches for interoperability of data streams).

4.4. Reusability

Ownership of datasets will belong to project consortium after the project completion. Creative Commons licence CC BY-SA or CC BY will be used for any open datasets unless there are compelling reasons to select a more restricted type of CC-licence. Creative commons licences will by default also include a disclaimer of liability for the re-use of opened data. The availability of data is summarised in Table 3.

No definite period or time limit is planned for access or re-use of the data. Justification for possible case-specific embargo for published data will be decided by the project consortium (steering board). Embargo will be sought primarily in connection with any potential patent application based on project results. The general rule follows what stated and agreed in the CA.

For all public open data, it will remain reusable via Zenodo for at least 20 years. As stated by the Zenodo Repository: “items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least”.

Table 3: Data availability and re-use in Lasers4MaaS

| Data | Type | Availability/re-use |
|--|--------------|---|
| Underlying data published in scientific journals | Public | Available after publication |
| | Confidential | Not available for publication because of confidentiality concerns |
| Data from public deliverables | Public | Available after approval of EC project officer |
| Other data | Public | Available after consent of all project partners at the end of the project at latest |
| | Confidential | Not available for publication and re-use |

4.5. Allocation of resources

The project manager and the project coordinator (WMG) will be responsible for updating the DMP throughout this project. Further they will provide guidelines to partners in the consortium in terms of compliance with the DMP. However, each partner has the responsibility of ensuring quality assurance in their data and metadata. The activities related to the data generation and collection are part of the person months (under the Lump Sum financial model of Lasers4MaaS) quantified in the GA for each partner.

The cost related to long-term preservation (and protection, if necessary) of the generated data is not quantified yet in this initial DMP, and depends, for example, on the data storage repositories used and length of time for which the data shall remain accessible. The costs corresponding to the open access provision of research publications and research data have been included as “Other goods and services” in the budget. The long-term preservation of generated data shall be reported in subsequent DMP versions, as we need to first obtain the total cost quantification before that. A detailed estimate will be available once the consortium has fully discussed the factors involved.

4.6. Data security

Lasers4MaaS will adopt a system that ensures reliable storage systems of data. Features will include automatic data backup procedures, controlled access mode even with collaboration (login required to access data), access to file version history, file naming and folder structuring. Open data, if any, will be stored in a trusted repository (Zenodo). Internally shared data are stored on MS TEAMS, managed by the project coordinator (WMG). Additionally, for data not directly shared (for example, raw experimental/simulation data) it is each partner's responsibility to host that data safely and redundantly.

5. Ethics

The project is expected to have a major positive impact on the environment leading to clean and climate neutral value chains and to a more circular economy through the digitalisation of the industry. The consortium sees no negative impact of the proposed technology on ethical dimensions as well as society, environment and laws.

The project complies with the “Do No Significant Harm” principle since its results will help to accelerate the adoption of decentralised, sustainable and circular production, alongside adoption of digital technologies. The project will help shaping the formulation of the digital product passport, backed by data interoperability. The consortium acknowledges the EU “Ethics Guidelines for Trustworthy AI” and its principle for lawful, ethical & robust implementation of AI. Our proposal will use AI technology to improve responsiveness and, therefore respond autonomously to changes in manufacturing systems. This will drive towards zero-defect and zero-waste. We have identified in the project the need to consider several aspects associated with the use and development of AI technologies with a holistic and global view: legal (compliance with safety requirements, etc.), social (public awareness, acceptability, trust, etc.), economic (reduce operational costs, add value for industry, etc.) and environmental impact.

We will consider transparency and traceability for trustworthy of ML/AI, related to the need to maintain a complete account of the provenance of data, processes, and artifacts involved in the production of the models. A highly relevant strength of Lasers4MaaS will be the use of explainable AI and interpretable ML to augment the capabilities of the analysis and control, which is critical to ensure robust and trustworthy AI. The consortium sees no negative impact of the proposed technology on ethical dimensions as well as society, environment and laws.