

Deliverable D6.1

Data Management Plan

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. Neither the European Union nor the granting authority can be held responsible for them.

Action number: #101138859

Action Acronym: DILAPRO
Action Name: Digital Laser Production

Deliverable Due date: 30-06-2024
Deliverable hand-in date: 18-06-2024

Revision	Date	Author/Organisation	Description
1.0	25.04.2024	DTI	Initial version
1.1	26.04.2024	DTU/DCU/IMDEA	Incorporating inputs of DTU, DCU, IMDEA
1.2	29.04.2024	DTI	DTI: Polishing of received inputs into a Q&A fashion
1.3	27.05.2024	PEPITe	Incorporated PEPITe comments
1.4	29.05.2024	DTI	Final version for proofreading
1.4	18.06.2024	DTI	Final version

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

Table of contents

Executive Summary **3**

Table of Acronyms **4**

1. Introduction **5**

2. Data Summary **7**

3. FAIR data **10**

3.1 Making data findable, including provisions for metadata 10

3.2 Making data accessible 11

3.3 Making data interoperable 14

3.4 Increase data re-use 14

3.5 Other research outputs 15

3.6 Allocation of resources 16

3.7 Data security 17

3.8 Ethics 17

3.9 Other issues 18

4. Conclusion **18**

5. References **19**

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

Executive Summary

The Data Management Plan (DMP) for the DILAPRO project outlines the data management strategies across the project, focusing on the main data generating partners DTU, DCU and IMDEA. PEPITe's DATAmaestro software will serve as a central repository for data collected from all partners, and their strategies for data handling in this system are also included in this DMP.

In the DILAPRO project, data is initially stored internally at the generating partner for collection and validation. It is then uploaded to the DATAmaestro Lake, where it is structured and combined with existing data. All data, including sensitive and imperfect datasets, is uploaded with different restrictions based on access requirements. Relevant datasets will also be uploaded to Zenodo for public access.

The DILAPRO project will adhere to the FAIR data policy, which aims to make data Findable, Accessible, Interoperable, and Reusable. Expected experimental datasets and modelling datasets have been identified, and metadata and identifiers will be provided for each dataset to ensure easy discoverability. Using the pilot project EU Open Research Repository at Zenodo for publication of relevant datasets will ensure the best possible compliance with the FAIR policy.

This DMP is a living document and will be updated throughout the project as new information becomes available. It sets the standard for data handling for all DILAPRO project partners and will be re-submitted as a deliverable in Month 30 with a focus on long-term storage and FAIR policy compliance.

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

Table of Acronyms

AM	Additive Manufacturing
API	Application Programming Interface
CAD	Computer Aided Design
DCMES	Dublin Core Metadata Element Set
DCU	Dublin City University
DILAFACT	Digital Laser Factory
DILAPRO	Digital Laser Production
DMP	Data Management Plan
DoA	Description of Action
DTI	Danish Technological Institute
DTU	Danish Technological University
EBSD	Electron Backscatter Diffraction
EDAX	Energy-dispersive X-ray analysis
EDS	Energy dispersive X-ray spectroscopy
FAIR	Findable, Accessible, Interoperable, Re-usable
GB	Gigabyte
GND	Gemeinsame Normdatai
HAZ	Heat Affected Zones
IMDEA	Institute for Materials and Advanced Studies
ISNI	International Standard Name Identifier
OPC-UA	Open Platform Communications Unified Architecture
ORCID	Open Researcher and Contributor ID
RAID	Redundant Array of Independent Disks
ROR	Research Organization Registry
RPC	Remote Procedure Call
SEM	Scanning Electron Microscopy
TB	Terabyte
TEM	Transmission Electron Microscopy

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

1. Introduction

This Data Management Plan (DMP) founds the basis for the DILAPRO Data management. It will commence with a general description of the main data generating partners, and the overall data management structure of the project. Next, the questions from the DMP template, written in *blue italic,* will be answered one by one. For each question, there will be either a description from the point of the project or specific descriptions for each of the main data generating partners, if the question prompts different answers across the partners. Descriptions from specific partners will be marked clearly with the bold and underlined shorthand name, i.e. **DCU**.

DILAPRO is a data-heavy project, and as such the Data Management Plan is of serious concern. All partners will generate data in one way or another, ranging from simple data collection from the printing process for all Pilot Line owners to large-scale numerical models or algorithms. Therefore, it has been decided to focus this initial DMP on the partners planned to do major data generation, but to involve all partners in the directions. The partners with major data generation have been determined to be DTU, DCU and IMDEA. Furthermore, PEPITe is responsible for the software called DATAmaestro, which will be used to collect the data from all partners and form the basis for the DILAFACT and DILACERT software. They will also generate some models through their DATAmaestro platform, however, as this is a part of a commercial solution this data will not be publicly available and will not be the main focus of this DMP. Their responses will also be included directly, when deemed necessary, and will in that case only refer to the data handling on the DATAmaestro Lake.

Below, a short introduction to each of these partners and can be found.

DTU will be responsible for developing multi-physics models of different aspects of the Additive Manufacturing process. They will re-use already existing models from other projects and generate new models on the areas that are lacking. They will use a combination of physical measurements to compare and validate the models.

DCU will develop process monitoring for both a laser powder bed fusion system and a laser surface treatment system. Data will be generated from in-process monitoring and post-process characterisation, to allow the development of accurate physical simulations of the process and the training of machine learning/artificial intelligence algorithms for prediction, real-time feedback, and quality control.

IMDEA will generate data from physical simulation experiments, using the GLEEBLE system to simulate the environment inside an Additive Manufacturing machine. Data will consist of a combination of data directly from the GLEEBLE system combined with data from external testing of the samples.

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

PEPITe's DATAmaestro software will be used as a joint repository for all DILAPRO data. Deliverable 1.3 (Description of relevant and current software workflows for the respective use cases) gives a good introduction to the 4 parts of the DATAmaestro package, which consists of: Collectors, Lake, Analytics and Dashboard. For the DMP, the DATAmaestro Lake will be most important, as it will be the main place for collecting all data to be used in the project. The purpose of DATAmaestro Lake is to store, consolidate and centralize data streams, and to resample, and prepare clean data tables for Advanced Analytics.

The data flow in the DILAPRO project will now be described. Initially, data will be stored internally at the generating partner for collection and validation. All usable data will be uploaded to the DATAmaestro lake, where it will be structured and combined with existing data both from DILAPRO and other projects. Here, data will be uploaded with different restrictions depending on who should be able to access it. All data will be uploaded, including sensitive datasets and imperfect datasets that can still be used to strengthen the models. The relevant datasets shall also be uploaded to Zenodo to allow for public access. Specifically, data will be uploaded to the community of Zenodo called the Open Repository for EU Research [1], which is a pilot project that will open in the autumn of 2024. Early access has been granted for the DILAPRO project, creating a collected community for upload of datasets and articles related to the DILAPRO project.

Furthermore, DILAPRO will ensure that all electronic files follow the FAIR policy. More in detail, the project partners have identified the typical datasets expected to be produced during the project. A summary of expected experimental datasets is provided in Table 1, while Table 2 outlines expected datasets from modelling. These lists are indicative and will be adjusted throughout the project by the addition, removal, or modification of datasets. Rich metadata for each experimental dataset will be provided to ensure that users can easily discover relevant datasets through various search criteria.

This DMP is an initial plan, which will set the tone for data handling for all the partners in the DILAPRO project. It will remain a living document to be updated throughout the project when new information comes into light. It will be re-submitted as a deliverable in M30, where more detailed explanations based on experience will be available. At M30, the focus will mainly be on the long-term storage to make sure the data complies with FAIR policy long after the DILAPRO project has ended.

Copyright © by DILAPRO All Rights Reserved

2. Data Summary

Please see the tables below [Table 1&2] for a general overview.

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

We will re-use computational materials data from the Materials Project database, the NOMAD database, and possibly other repositories to obtain relevant thermodynamic parameters for establishing relevant numerical models. Existing validated open datasets from repositories such as Zenodo will also be re-used for reducing the number of experimental campaigns in DILAPRO.

What types and formats of data will the project generate or re-use?

DILAPRO will encompass a large number of different materials characterization techniques. In Table 1, an overview of different formats of data is presented.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

DCU: Monitoring and characterisation data, which are generated through experiments, serve the objectives of the project by providing training and validation data for machine learning algorithms and physical simulations (incl. set up of a recommender system).

DTU: The generation of process simulation data helps the generation of established machine learning models, which will also generate extensive relational datasets. Process simulations will also be used as complementary method of generating material properties data in lieu of extensive experimentation. The simulation data will also be utilised to map process parameter windows and stable process signature ranges to meet DILAPRO use-case specifications.

IMDEA: Data generation creates new, specific data to directly answer questions related to the microstructure evolution during processing and mechanical performance of the final products.

What is the expected size of the data that you intend to generate or re-use?

DCU: The expected total data size for laser processing could be in the GB range, while for the metal AM, this would be in TB range, in the lifetime of the project.

DTU: Raw (unprocessed) data sets from simulations will span 20-40 TB over the project period, while post-processed data for integration with the DILAFACT data warehouse is expected to be generated at 7-10 TB/year.

IMDEA: The size of the data generated during this project will be of a very wide range (from kB to several GB).

Copyright © by DILAPRO All Rights Reserved

What is the origin/provenance of the data, either generated or re-used?

DCU: In the laser processing and metal additive manufacturing use-cases, monitoring and characterisation data will be generated through experiments.

DTU: Finite element and finite volume simulation data - in hdf5 or similar formats, will be generated during numerical modelling of the DILAPRO processes. Also, reduced order models for laser-material interaction, and the resulting defects, microstructural features and stresses, will be produced from the same source. Correspondingly from these reduced order models, further process simulation data will also be generated.

IMDEA: The experimental data for the project will originate from two sources. Firstly, internally generated data will stem from experimental work and advanced characterization conducted within the project. In particular, the experimental data will be obtained from software installed on characterization equipment (e.g. microscopes for microstructural analysis, GLEEBLE system, facilities for mechanical testing). Secondly, external data will be re-used for speeding up the research activities (such as drawings of sample geometries, etc.) or for benchmarking of project results (e.g., comparing the project results with those for similar materials).

To whom might your data be useful ('data utility'), outside your project?

The post-processed data sets will be made accessible through data repositories, and multiple other R&D projects around laser metal processing will benefit from this. Furthermore, the datasets may be used or referred later for parts reproduction and/or process optimisation, or for materials scientists working on metallurgy, microstructural and mechanical characterization of steels, Ni-based superalloys and Al alloys.

Table 1. Summary table of the description of expected datasets from experimental studies in the DILAPRO project.

Data type	Data format	Data size	Data origin	Partners producing data	Re-use existing data
Optical microscopy images	.tiff, .jpg, .png	~100 kB – ~10 MB	ANALYSIS software,	IMDEA, DCU, DTU, tba	No
Scanning electron microscopy images	.tiff, .jpg, .png	~1 MB – ~10 MB	SEM software,	IMDEA, DCU, DTU, tba	No
Transmission electron microscopy images	.tiff, .jpg, .png	~1 MB – ~10 MB	TEM software, EDS: Analysis station	IMDEA, tba	No
EBSD data	.ctf, .crc, .cpr, .h5oina, .h5, hdf5, .ang, .osc, .oim, .png, .jpg, .txt	~10 MB – ~1 GB	AZTEC software, EDAX: TSL OIM	IMDEA, tba	No

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

CAD models of samples for testing	.dwg, .pdf	~10 kB – ~100 kB	AutoCAD, SolidWorks, CATIA, SpaceClaim, NetFabb	IMDEA, DCU, DTU, tba	Yes
Data from tensile testing	.xlsx, .csv	~100 kB – ~1 MB	INSTRON software (or similar)	IMDEA, DCU, tba	Yes
Data from high cycle fatigue testing	.xlsx, .csv	~100 kB – ~1 MB	INSTRON software (or similar)	IMDEA, tba	Yes
Data from fracture testing	.xlsx, .csv	~100 kB – ~1 MB	INSTRON software (or similar)	IMDEA, tba	Yes
Data from GLEEBLE testing	.xlsx	~100 kB – ~1 MB	INSTRON software (or similar)	IMDEA	Yes
Data from hardness testing	.xlsx, .png	~1 MB – ~20 MB	Hardness tester software	IMDEA, DCU, tba	Yes
Laser process monitoring data	.csv, .pcd	~10 kB - ~300 GB	Acoustic, pyrometer, and eddy current sensors	DCU	No
Data from surface profiling	.OPDx	~1 MB - ~10 MB	Bruker contour non-contact profilometer	DCU	Yes
Data from contact angle measurement	.jpg, .mdb	~1 MB - ~100 MB	FTA200 Dynamic Contact Angle Analyser	DCU	Yes
Data from 4D-Xray characterization of samples during in-situ heat treatment	.csv, .hdf5	~100GB-~1TB	In-house 4D-Xray characterization setup	DTU	No
…	…	…	…	…	…

Table 2. Summary table of the description of expected datasets from modelling in the project.

Data type	Data format	Data size	Data origin	Partners producing data	Re-use existing data
Data from simulation of HAZ thermal cycles	.xlxs	~10 kB – ~100 kB	Excel	IMDEA	No
Data from process Multiphysics simulations	.hdf5, .vtk, .rth, .csv,. odf	~100GB-~10TB	Numerical models implemented in in-house /commercial software	DTU, DCU	No
Data from machine learning models	.csv,.hdf5	~100GB-~1TB	Machine learning models	DTU, DCU, PEPITe	No
…	…	…	…	…	…

3. FAIR data

3.1 Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

PEPITe: Every time series is uniquely identified via a virtual path; i.e. one or multiple directories and a name, based on the identifiers from each of the partners.

DCU: Data associated with the AM and laser processing at DCU will be identified using a metadata system on an internal data storage server.

DTU: Data associated with the DILAPRO simulations and testing at DTU will be identified using a metadata system on an internal data repository (DTU Data) hosted by DTULibrary.

IMDEA: Data associated with the physical simulation experiments, microstructure and mechanical characterization will be identified using a metadata system on an internal data server.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

The metadata will follow the Dublin Core Metadata Element Set (DCMES) [2] standard or similar. The Dublin Core metadata element set is a standard for cross-domain information resource description, providing a simple and standardized set of conventions for describing things online in ways that make them more discoverable. In DILAPRO, at minimum these descriptions will be used: data owner name, material, application, process parameters, date, contributors, and characterisation types.

In the DATAmaestro Lake, the metadata is at minimum the data type. Any additional metadata can be added by the users if it is relevant to their discipline via a system of classifiers (e.g. country=be ; where country is a classifier and be a value for this classifier).

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

The metadata will be searchable by (where applicable): data owner name, material, process parameters, application, characterization type (i.e. optical microscopy, fatigue testing), characterization parameters, date, contributors, and characterisation type.

This metadata shall be available both internally, in the DATAmaestro Lake and at the Zenodo repository and can be re-used.

Copyright © by DILAPRO All Rights Reserved

Will metadata be offered in such a way that it can be harvested and indexed?

A harvestable data standard such as the Dublin Core Metadata Element Set (DCMES) will be followed for the storing of all metadata, allowing the data to be easily parsed for indexing and harvesting. In DATAmaestro Lake all metadata can be retrieved via the web API provided the credentials has authorization to access the data.

3.2 Making data accessible

Repository: Will the data be deposited in a trusted repository?

Initially, data will be stored internally on site in a data bank/server at the generating partner. All usable data will be uploaded to the DATAmaestro lake, where it will be structured and combined with existing data both from DILAPRO and other projects. The relevant datasets shall also be uploaded to Zenodo at the EU Open Research Repository.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

On-site repositories for all partners have been arranged. Early access to the EU Open Research Repository has been granted through the page https://zenodo.org/communities/dilapro/records.

DTU: Availability and appropriateness of using a centralized national repository (DeiC in Denmark) will be investigated during the project period.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

PEPITe: Every time series is uniquely identified via a virtual path; i.e. one or multiple directories and a name.

Data: Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Data will be made openly available where suitable. Access may be restricted on commercial grounds or to allow the researchers who generated the data to make use of it before it is made public.

PEPITe: Data will not be made openly available on the DATAmaestro Lake, but only on Zenodo. Credentials and associated authorization will be required to access any data.

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Access will be restricted initially, to allow the researcher who produced each specific dataset to pursue publication, with batches of data being shared with project partners case-by-case. Data associated with publications will be attached as supplementary data as needed, as brief and concise as possible. The typical time from data generation to full open access will therefore span from 1-3 years.

Will the data be accessible through a free and standardized access protocol?

An open, standardised metadata protocol, DCMES (ISO 15836) [2], will be used to organise the data to be accessible, navigable, and harvestable.

PEPITe: The data will be available through a Remote Procedure Call (RPC) web API. There is also a web API supporting OData protocol. Data can also be downloaded in open format such as .CSV or Numpy.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

Access may be restricted to allow primary use of data by the generating party or to protect commercial interest, with access being mediated on a case-by-case basis. Data will not be uploaded for public availability before being cleared by the generating party. During and after the project all data will be available on the DATAmaestro Lake upon request, although access will be denied for commercially restricted datasets.

How will the identity of the person accessing the data be ascertained?

PEPITe: Data access to the DATAmaestro Lake will be granted by validating user credentials and associated authorization. Authentication can be handled by any provider supporting OpenID (OAuth 2).

Data access on Zenodo or EU Open Research Repositories will follow their guidelines.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

A data access committee may be determined from DILAPRO staff and industrial partners where applicable, to determine appropriate restrictions and decide on release of restricted data.

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

Metadata: Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

The records in the EU Open Research Repository are required to comply with the following minimal metadata requirements: [3]

- Visibility: Both public and restricted (with or without embargo and/or access request)
- Resource types: All resource types.
- Licenses: Public and embargoed records MUST specify a license.
- Funding information: Records MUST specify at least one grant from the European Commission.
- Creators: Creators SHOULD be identified with a persistent identifier (e.g. ORCID, GND, …), and affiliations SHOULD be identified with a persistent identifier (e.g. ROR, ISNI, …)
- Subjects: Records SHOULD specify one or more fields of science from the European Science Vocabulary.

Additional metadata will be made openly available unless restricted by commercial interest.

PEPITe: Metadata will be available to authenticated user on DATAmaestro Lake. Data and metadata access are given together.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Data and metadata will remain on Zenodo through the EU Open Research Repository for as long the platform allows, except if some datasets become restricted after initial publication.

PEPITe: Data and metadata will remain available on the DATAmaestro platform as long as the projects requires it. Data and metadata are deleted together.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

SOPs and data documentation will be stored in the repositories and linked to the datasets via the metadata system.

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

3.3 Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

Characterisation and monitoring data will be stored in open formats where possible, using standardised metadata. The project will ensure that all electronic files follow the FAIR policy.

PEPITe: Data can be retrieved from DATAmaestro Lake as .CSV or Numpy file. The web API makes it available via HTTP where the data is returned as .JSON format. The data can also be pushed back to any platform supporting an OPC-UA interface.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Specific ontologies and vocabularies are dependent on the projects and can be edited at will in the DATAmaestro system. Acronyms or specific vocabulary will be defined at first use in any documentation or publication.

Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

Data will include qualified references to other data and/or metadata from previous research, particularly for benchmarking.

3.4 Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Supplementary information such as experimental or data processing procedures will be stored on the internal servers and connected to the relevant data using the metadata system.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

All usable data (including imperfect datasets and commercially sensitive data) will be uploaded to the DATAmaestro lake, where data will not be made freely available. Credentials and associated authorization will be required to access any data. However, all relevant datasets will also be uploaded to Zenodo through the EU Open Research Repository to allow for free public access to allow for the widest re-use possible. Public access may be restricted initially to allow publication or out of IP considerations. Data will be licensed using standard reuse licenses, and

Copyright © by DILAPRO All Rights Reserved

the project will investigate the use of the REUSE software to assist in this matter (https://reuse.software/).

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Any generated data will be re-useable, within restrictions on the basis of academic or commercial interests. The data will remain on the public server (Zenodo) for an indefinite period of time.

Will the provenance of the data be thoroughly documented using the appropriate standards?

Data provenance can be documented via classifiers.

Describe all relevant data quality assurance processes.

- Using open standard data formats as much as possible (HTTPS, JSON, OPC-UA, Odata, Oauth2, ...).
- Automating data collection
- Strong typing
- Data collection monitoring
- Backups

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

3.5 Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

Samples will be labelled and stored on-site after characterization to allow retrieval. Software will be developed using open tools such as Python to enable interoperability. Digital Twins will stay in open source or educational licenses to permit the development of modules based on

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

new materials and sensors. These modules then can be integrated in the DILAFACT environment with royalties or specific agreements for the developers.

3.6 Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

The cost for making articles Open Access may be up to 2-4000 € per article. DILAPRO has a Key Performance Indicator of publishing at least 5 scientific articles during or immediately after the project. No other costs are foreseen at this stage.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

6000 € is reserved in the DILAPRO project for providing Open Access to research articles to be published by IMDEA. This was not included in the budgets for the remaining partners, who will incur the costs themselves.

Who will be responsible for data management in your project?

Danish Technological Institute as the coordinator will be responsible for the overall plan for data management. The individual partners will be responsible for adhering to the plan to correctly manage the data throughout the project.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Long term preservation will be ensured by using a combination of DATAmaestro and Zenodo, along with the internal on-site servers. Data will as a general rule be preserved for as long as the platform allows (Currently there is no limit), but individual datasets can be removed if deemed necessary for any reason.

PEPITe: For the DATAmaestro software, data is stored on disk on ext-4 filesystem; it can be locally on-site or on a cloud provider of choice. The size can be extended if it is required, and the cost depends on the underlying infrastructure. The required storage space can be estimated by the following hypothesis: 16 bytes per data point (e.g. 2000 tags every 15 seconds: 2000 * 365 * 24 * 60 * (60 / 15) * 16 bytes = 67.3 GB / year).

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

3.7 Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

In terms of the project structure, the data will be located both internally, in the DATAmaestro lake and in some instances at the Zenodo. This creates a solid route for data recovery, along with the internal systems described below.

PEPITe: Daily backup with snapshots, encryption end to end for data in transit, encryption at rest for stored data. The storage can on local hardware or cloud storage depending on requirement.

DCU: The local data storage server in DCU uses a RAID system to allow for full data recovery in the event of a hard drive failure. Data will also be backed up on PEPITe's DATAmaestro software.

IMDEA: We ensure that our data is backed up regularly on a secure server. We use passwords, firewalls, and antivirus software to protect our data from unauthorized access or loss.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Data will be stored both internally, in the DATAmaestro lake and for publicly accessible data at the Zenodo for long term preservation.

3.8 Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

The data present no specific ethical or legal issues. The sharing of personal data may be an ethical concern. The only personal data expected will be researcher names in the metadata to ensure credit and attribution, for which consent will be obtained from the researcher. Access to data will initially be restricted to allow the generating individuals to make use of it before it is shared. Any ethical concerns or issues that arise will be raised with DTI, as laid out in Task 6.3 of the project document.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

Personal data (researcher names in metadata) will be gathered with informed consent from the researchers.

Copyright © by DILAPRO All Rights Reserved | D6.1 – Data Management Plan

3.9 Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

We do not plan other procedures for Data Management at this stage.

4. Conclusion

In conclusion, this DMP serves as a robust guide for data generation, storage, and re-use across participating partners. It outlines a structured pathway for data management, ensuring adherence to the FAIR policy, thus promoting data discoverability and usability.

The DMP is not a static document but will evolve as the project progresses, incorporating new insights and experiences. The next update to this document will be in month 30, where focus will be on securing long-term storage to ensure compliance with the FAIR policy even after completion of the project.

This approach to data management underscores the DILAPRO project's commitment to transparency, collaboration, and the advancement of knowledge in the field of Laser Processing.

Copyright © by DILAPRO All Rights Reserved

5. References

[1] [Online]. Accessed 18-06-2024. Available: https://zenodo.org/communities/eu/.

[2] ISO 15836: Information and documentation — The Dublin Core metadata element set.

[3] [Online]. Accessed 18-06-2024. Available: https://zenodo.org/communities/eu/curation-policy.

[4] [Online]. Accessed 18-06-2024. Available: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/.

Copyright © by DILAPRO All Rights Reserved