

Grant Agreement number: 871161

Project acronym: **IMPULSE**

Project title: '**Integrated Management and reliable oPerations for User-based Laser Scientific Excellence**'

PROJECT DELIVERABLE 1.4

PROJECT DATA MANAGEMENT PLAN

Deliverable leader:	ELI Delivery Consortium (ELI-DC)
Authors:	Florian Gliksohn, Teodor Ivařnoaica, Allen Weeks
Due date:	30 April 2021
Actual submission date:	9 May 2021
Dissemination level:	<input type="checkbox"/> Confidential, only for members of the consortium (including the Commission Services) <input checked="" type="checkbox"/> Public
Abstract:	
<p>The Data Management Plan (DMP) presents and supports the data management life cycle for all data that will be collected, processed, or generated within the framework and in relation with the IMPULSE project. It provides a description of the data types the project will generate, of the way the data will be collected and stored and made available for validation, exploitation and re-used at the partner organisations. The DMP is, by nature, a living document that will be updated when needed during the project.</p>	



Document Revision History:

Date	Version	Author/Editor/Contributor	Summary of main changes
08/04/2021	V1	Teodor Ivănoaica	First consolidated version
15/04/2021	V2	Florian Gliksohn	Review and submission to Steering Board
23/04/2021	V3	Teodor Ivănoaica	Second revision
30/04/2021	V4	Florian Gliksohn	Review and submission to Steering Board
09/05/2021	V5	Florian Gliksohn, Allen Weeks	Final version and submission to EC
10/05/2021	V6	Teodor Ivănoaica	Third Revision

Disclaimer

The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Commission. The European Commission is not responsible for any use that may be made of the information contained therein.

Copyright

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the IMPULSE Consortium. In addition, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

Contact



In case of any questions or clarifications regarding this deliverable, contact the Project Management Team at the following address: pmt-impulse@eli-laser.eu.



TABLE OF CONTENTS

I.	Introduction	6
1.1	Purpose	6
1.2	Relevant provisions of the Grant Agreement	6
II.	Data Summary	7
III.	Document Repositories	10
IV.	Fair data generated within IMPULSE	10
V.	Tools to access data	11
VI.	Costs	12
VII.	Data Privacy and Security	12



LIST OF ABBREVIATIONS

Abbreviation	Meaning
DMP	Data Management Plan
EC	European Commission
ELI	Extreme Light Infrastructure
ELI ERIC	ELI European Research Infrastructure Consortium
ERIC	European Research Infrastructure Consortium
GA	Grant Agreement
GDPR	General Data Protection Regulation (EU) 2016/679
ORD pilot	Open Research Data Pilot
WP	Work Packages



I. INTRODUCTION

1.1 Purpose

The purpose of this Data Management Plan (DMP) is to present and support the data management life cycle for all data that will be collected, processed, or generated within the framework and in relation with the IMPULSE project. It provides a description of the data types the project will generate, of the way the data will be collected and stored and made available for validation, exploitation and re-used at the partner organisations.

The project DMP is, by nature, a living document that will be updated when needed during the project. In particular, the ELI ERIC Data Policy is expected to be adopted in one of the initial meetings of the ELI ERIC General Assembly; the DMP will be updated to ensure full consistency with the Data Policy.

1.2 Relevant provisions of the Grant Agreement

IMPULSE is part of the pilot launched by the European Commission (EC) under Horizon 2020 called the 'Open Research Data Pilot' (ORD pilot). The ORD pilot aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects.

As per Article 29.3 of the IMPULSE Grant Agreement (GA), the following obligations apply to the IMPULSE beneficiaries within the framework of the ORD pilot. The beneficiaries have the obligation to "deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

- the data, including associated metadata, needed to validate the results presented in scientific publications, as soon as possible;
- other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan'."

Beneficiaries shall also "provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves)."

The ORD pilot deals with *datasets*, not with scientific peer-reviewed publications for which open access is an obligation as per Article 29.2 of the GA. The latter provision sets forth the obligation for each beneficiary to "ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results". Beneficiaries must:

- "as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications"



- “ensure open access to the deposited publication — via the repository — at the latest: (i) on publication, if an electronic version is available for free via the publisher, or (ii) within six months of publication
- ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication.
- The bibliographic metadata must be in a standard format and must include all of the following: the terms “European Union (EU)” and “Horizon 2020”, the name “Integrated Management and reliable oPerations for User-based Laser Scientific Excellence”, the acronym IMPULSE and grant number 871161, the publication date, and length of embargo period if applicable, and a persistent identifier.

II. DATA SUMMARY

The IMPULSE project will generate two major categories of Data:

- The first category includes documents relates to the management of the project (e.g., meeting minutes, reports, deliverables, concept documents), ELI management processes and workflows and software designed to serve the Integrated Data Management System of the ELI ERIC
- The second category includes all scientific data generated mainly within the framework of the access pilots of tasks 5.3, 5.4 and 5.5 Work Package that will be governed by the FAIR principles.

Data belonging to the first category will be made available mainly via the ELI Confluence page. The public documents and deliverables will be also made available via the impulse project website. The table below provides a detailed overview of the data types for this first category.

Data Type	Contents	Data Size / Volume	Archive
Text files, Word documents, PDF documents	Minutes, technical documentation, deliverables, milestones, reports	100s of Megabytes	Confluence IMPULSE website
Text files, Text Documents, Charts, Surveys, Forms	Surveys, charts generated based on collected feedback and documents shared for collaborative work.	10s of Megabytes	Collected via Google Workspace. Archive: Confluence
Diagrams, charts, workflows	Technical documentation, management system documentation, work plans	100s of Megabytes	Confluence IMPULSE website
Source Code	Source code developed in the development/deployment/implement	10s of Megabytes	Github.com



	ation process		
Training materials	Training materials (publicly available or not) including course description and course catalogues	10s of Gigabytes	ELI e-learning platform



Data generated within the framework of the access pilots performed at the three ELI Facilities will be treated as FAIR data and made available via the ELI Data Portal. The Data Portal instance, currently under development with the support of the PaNOSC project and IMPULSE, will be operated and maintained by ELI ERIC. Data will be searchable, accessible, and reusable via the Data Portal, each ELI Facility providing full details about the existing metadata and naming conventions so that the datasets are also made interoperable. The datasets of other research works derived from the project will be treated in compliance with Article 29.2 of the Grant Agreement.

Data Type Access pilot	Contents	Data Size / Volume	Archive	Origin	Format	Reusable
Journal articles	Articles disseminating results generated within the framework of the project	100s of Megabytes	Peer-reviewed open access journals, conference proceedings	Access pilots, other publications linked to WP3 and WP4	PDF	Open Data Policies
Scientific Data from access pilots	Raw Data, metadata and associated auxiliary data collected and derived from scientific experiments and laser/beamlines operations, documentation on metadata formats	100s of Gigabytes	ELI user portal, local storage + ELI ERIC archiving systems	Access Pilot, experimental devices...	HDF5 - data metadata - nexus (other?)	Local Beam scientists (it can also be NOT to be reused)
Simulated data	Simulated data sets resulting from tasks T5.3, T5.4 and T5.5 (T5.2, 5.6)	100s of Megabytes	ELI user portal, local storage + ELI ERIC archiving systems, Github.com			



III. DOCUMENT REPOSITORIES

IMPULSE will be using two main document repositories:

- General text documents will be stored also in the ELI ERIC Confluence spaces: <http://eli-eric.atlassian.net/>
- Text files associated with the software or documentation of the codes developed in the context of the IMPULSE PROJECT will be stored in the ELI ERIC GitHub: <https://github.com/eli-eric>.

For each work package, a dedicate space has been created in Confluence:

- WP1 (**Setting-up and project management**) <https://eli-eric.atlassian.net/wiki/spaces/WP1/overview>
- WP2 (**Joint development of ELI as an integrated Organisation**) <https://eli-eric.atlassian.net/wiki/spaces/IW/overview>
- WP 3 (**Ramping-up Towards Excellent Steady-State Operations**) <https://eli-eric.atlassian.net/wiki/spaces/WRTESO/overview>
- WP4 (**Key technologies and enhanced experiments**) <https://eli-eric.atlassian.net/wiki/spaces/WKTEE/overview>
- WP5 (**Enabling excellent user access**) <https://eli-eric.atlassian.net/wiki/spaces/WEEUA/overview>
- WP6 (**Fostering ELI's Innovation Impact**) <https://eli-eric.atlassian.net/wiki/spaces/WFEII/overview>
- WP7 (**Promoting ELI Membership & Communication**) <https://eli-eric.atlassian.net/wiki/spaces/IW7/overview>

The expected data volumes for the various types of documents and source codes developed in IMPULSE are expected to range from 10s to 100s of Megabytes.

IV. FAIR DATA GENERATED WITHIN IMPULSE

The FAIR principles will be applied only for scientific data generated by the IMPULSE project activities. The tasks focusing on the optimisation of the performance of the ELI technology and experimental systems and the access pilots, planned within the framework of WP5, will be the sources of scientific data that will be managed and made available following the FAIR data principles.



All partners will use the Nexus Metadata standards and/or provide a full documentation of the metadata standard that was selected. All metadata will be ingested in the ELI ERIC file catalogue and made available via the data portal.

For each Access Pilot, the task leader will provide information confirming that the following requirements are effectively fulfilled:

- The data produced and/or used in the project should be discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)
- A naming convention should be available
- Search keywords should be provided to optimize possibilities for re-use
- Clear version numbers should be provided
- In case metadata standards are not implemented, the type of metadata that will be created and the method of creation should be outlined.

Scientific data generated by IMPULSE will be searchable via the Data Portal catalogue (or temporary systems used before the catalogue is in place) or by using the search features on GitHub for the data stored there.

For the software solutions used and developed within IMPULSE, open-source software will be mainly used, with the source code and associated documentation made available via GitHub. Well-established open-source licenses such as Create Commons, Apache2 or MIT will be used. In rare cases where teams will develop or co-develop software using licensed software or using third-party licensed tools / service / software, that software will not be made available. For some of the situations, providing the machine code will be considered.

All IMPULSE scientific data and software source code will be organized in a logical way, making them easily browsable. GitHub has a search facility for software stored there.

The documents that are restricted to members only will be stored in Confluence respecting the necessary privacy policies.

V. TOOLS TO ACCESS DATA

The following tools and software will be required to access the data:

- Text documents generated by the IMPULSE Project will require a basic text editor
- Deliverables will require PDF reader to read (a pdf preview reader is embedded in Confluence)
- Confluence access is provided to all project members and is including support for online editing plugins.



Open Source provides anonymous access for reading. Identity will be managed via GitHub, authentication mechanisms for modifying source code and editing documents on GitHub. Anonymous access will be considered also for Open Data.

As the project evolves, parts of the documents generated by the project will be made available based on an Umbrella ID, which is a well-known Identity provider for Photon and Neutron Scientific Community.

VI. COSTS

The following costs are involved in making the data available:

- Licensing fees related to the Confluence user licenses provided for the members of the IMPULSE Project under the ELI-DC Community license
- Documents and source code are hosted in GitHub for free, The ELI ERIC GitHub organisation will be used for this project, WP Leaders/Task Leaders will provide a list of contributors
- Hosting fees for IMPULSE website.

Each WP leader is responsible for documents and source code management regarding his/her WP, with the Project Management Team overseeing the fulfilment of the data management responsibilities.

Resources for the long-term preservation of the data generated within the framework of the project will be provided by ELI ERIC in compliance with the requirements of the Grant Agreement.

VII. DATA PRIVACY AND SECURITY

Data privacy requirements will be fulfilled in compliance with the GDPR and principles and rules detailed in deliverable 8.1 – ‘POPD Requirement 1’.

