



XLS Deliverable D1.2

WP1: Data Management Plan (DMP)

Prepared by: Gerardo D'Auria, Evangelos Gazis,
Andrea Latina and Regina Rochow

Prepared on: 29-06-2018

This project is funded by the European Union's Horizon2020 research and innovation programme under Grant Agreement No. 777431.

The contents of this report reflect only the view of the CompactLight Consortium. The European Commission is not responsible for any use that may be made of the information it contains.

XLS | CompactLight¹

D1.2 - DATA MANAGEMENT PLAN (DMP)

Based on the EU DMP Template

Gerardo D'Auria, Evangelos Gazis, Andrea Latina, Regina Rochow

Introduction

This document is based on the Horizon 2020 DMP template, which is designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. It is envisaged to develop a single DMP for the CompactLight project to cover its overall approach.

FAIR data management

In general terms, our research data should be 'FAIR', that is Findable, Accessible, Interoperable and Re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

This document is not intended as a strict technical implementation of the FAIR principle; it is rather inspired by FAIR as a general concept.

Structure of the DMP

The document is a set of statements with a level of detail appropriate to the project.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project.

Rather, the DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

Therefore, the DMP of CompactLight has a clear version number and includes a timetable for updates: As a minimum, it is planned to update the DMP in the context of the Annual Review Meetings of the project. Since in CompactLight no other periodic reviews are envisaged within the Grant Agreement, an update will be made in time for the final Review at the latest.

In the following the main sections to be covered by the DMP are outlined. At the end of the document, Table 1 contains a summary of these elements in bullet form.

This template itself may be updated as the policy evolves.

¹ This project has received funding from the European Union's Horizon2020 Research and Innovation Programme under Grant Agreement No. 777431.

XLS | CompactLight

DATA MANAGEMENT PLAN Version 1.0

Project Number: 777431

Call: H2020-INFRADEV-2016-2017/H2020-INFRADEV-2017-1

Project Acronym: XLS

Project title: CompactLight

Date: 29 June 2018



1. Data Summary

The purpose of the data production, collection and processing is to gather all the scientific information of our XLS collaboration, which is defined as “open” by the partners, to be used subsequently by other researchers and promote the knowledge dissemination for the benefit of the scientific society.

There will be various types and formats of our project data to be collected. The collected data should be stored to one of the two categories: Scientific Data and Administrative Data.

The Scientific Data are:

- Calculations
- Simulations
- Measurements
- Designs
- Software
- Presentations
- Reports
- Documents

The Administrative Data are:

- Budget issues
- Transfer Technology issues
- Backup security & data protection
- Access to the data

Any existing data from the literature that will be re-used should be modified according to our data formats, taking into account and acknowledging that the origin of the data comes from the scientific and engineering work carried out in the frame of the XLS collaboration. It is expected that there will be no extremely big sized data sets, considering that the largest data sets will be the engineering design data. It is assumed that the future data of our collaboration can be extremely useful for research institutions and funding agencies aiming to build hard X-ray sources of the X-Free Electron Laser (XFEL) type with an innovative way of electron acceleration, developed from the CLIC² technology. As a result, our project will offer novel and innovative XFEL designs with compact, low-sized accelerators as well as low construction and operation cost.

2. FAIR Data

2. 1. Making data findable, including provisions for metadata

It is the aim of our collaboration that all the open data resulting from our project can be easily and simply found via a well-known list of keywords in the web. A common identification procedure will be applied via a digital object identification mechanism. The naming convention of our data will follow the names presented in our data categorization in section 1 - Data Summary.

User-friendly keywords will be implemented for our data search via the web. It should be noted that a clear version number will be provided for the DMP document and the standards of metadata will be mainly used for the engineering design data.

2.2. Making data openly accessible

Most of the Scientific Data will be openly available, at least those published in international scientific journals or presented in international conferences and published in their proceedings. Some of the

² CLIC stands for the Compact Linear Collider R&D project at CERN

sensitive Administrative Data cannot be available, i.e. detailed financial data concerning the budget of each participating institution. This restriction is also derived from the internal rules and regulations applied by each institution in those matters, as well as from current European law.

In the CompactLight project, data can be open or kept confidential for exploitation according to the provisions of Grant Agreement and Consortium Agreement. Open data will be accessible in a web-based repository and standard data base software tools available in the market for documents and designs will be implemented to facilitate the access to these data. A written introductory document will explain how to proceed in order to find the requested subject, using suitable software (e.g. in open source code). A certified repository owned by the CompactLight collaboration will host all kinds of open data and associated metadata, documentation, and code from the project for dissemination.

There will be no restrictions for access to open data, with the exception that any external scholar requesting the data should declare his name and affiliation. A Data Management Committee will be appointed in WP1, which will establish a procedure for granting access that permits to identify the requesting person (by sending an access code via email) and which will provide guidance for access to our data.

2.3. Making data interoperable

The data of our project will be interoperable, allowing data exchange and re-use between researchers, institutions, organizations, countries, etc. For that reason, there will be adhered standard formats as much as possible compliant with available open software applications and in particular facilitating re-combinations with different datasets from different projects. The terminology used for scientific data and metadata will follow standards and methodologies that allow to make our data interoperable. The applied terminology will be kept for all our data types in our data sets, allowing inter-disciplinary interoperability.

The project team will avoid using uncommon or newly generated project specific ontologies or vocabularies for the data of our project.

2.4. Increase data re-use (through clarifying licences)

In order to permit the widest possible re-use of the open data of CompactLight, as far as possible access to them will not be based on a licence. Additionally, most of the data of our project will be available immediately after approval by the researchers collaborating in our project. Any effort will be made to reduce the time for publication or patent request as much as possible. The data will be used or useable by third parties and in particular after the end of the project without restriction for re-use.

The duration of the data will be defined and the available means for the maintenance and the data quality assurance process will be described by the CompactLight collaboration in following versions of this DMP.

3. Allocation of Resources

There is yet no precise estimation of the costs for making the data of our project FAIR. A future cost evaluation will take into account a data base server with maintenance plus one FTE for the total period of the project duration (01/01/2018-31/12/2020).

The costs related to open access to our research data will be covered by the H2020 grant received for the XLS project. The responsible for the data management will be the XLS scientific coordinator.

It is envisaged to make the open data of the CompactLight project available for a long duration. The resources for long-term preservation will be discussed, agreed, and approved by the XLS collaboration.

4. Data Security

The data security process is split into two concepts. Security with respect to the storage media will be achieved through multi-type backup repositories, while data security with respect to data handling by users will be ensured through access limited to reading and downloading data sets without the possibility of changing / manipulating / saving them in the repository.

The data will be stored in safe repositories that enable long-term preservation, which will be selected by the project collaboration.

5. Ethical Aspects

Currently we do not expect any ethical or legal issue related to sharing XLS data openly. It will be ensured that the open data sets shared and preserved on the long term will not contain any personal data or other sensitive information, except for author names, contact data or other information expressly agreed with the concerned persons.

6. Other Issues

It is not foreseen to use any other national /funder /sectorial / departmental procedures for our data management.

7. Further Support in developing the DMP

The advice from the Research Data Alliance providing a Metadata Standards Directory that can be searched for discipline-specific standards and associated tools will be taken into account. The EUDAT B2SHARE tool includes a built-in license wizard, facilitating the selection of an adequate license for research data, if necessary.

The repositories will include listings of Administrative and Scientific Data, as classified in chapter 1. The research data repositories will be registered properly. The open data of the XLS project will be stored in the CERN repositories, which allow researchers to deposit both publications and data, while providing tools to link them. All the scientific tools for data management used in the domain of High Energy Physics will be implemented.

SUMMARY TABLE 1**FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP**

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
<p>1. Data Summary</p>	<ul style="list-style-type: none"> • The purpose of the data production, collection or processing is to gather all the scientific information to be used by next researchers and promote the knowledge dissemination for the benefit of the scientific society • The objectives of our project are briefly to deliver an innovative and low size, low cost X-free electron laser design, available to the scientific communities of our domains. • There will be various types and formats of project data from CompactLight to be collected. The collected data should be stored to one of the two categories: Scientific data and Administrative Data. • Any existing data should relate with the origin of the data. • The origin of the data may come from any scientific literature available to our domain. • The expected size of our data is not well known yet. • The data of our XLS collaboration are expected to be very useful to the scientific communities performing R&D based on light sources and particle beams using the proposed novel accelerator technologies, research institutions looking to build such facilities, funding agencies aiming to support financially the construction of such facilities.
<p>2. FAIR Data 2.1. Making data findable, including provisions for metadata</p>	<ul style="list-style-type: none"> • The best way for facilitating the future researcher to re-use and get profit of our data is to store the data with a simple and easy method including well-know keywords to be discovered and mined. • A common identification procedure will be applied via a digital object identification mechanism. • The naming convention of our data will follow the names presented to our data categorization in the 1 - Data Summary. • Friendly use keywords will be implemented for our data search via the web. • Clear version numbers will be provided to the DMP document. • The standards of metadata will be, mainly, used for the engineering design data.



<p>2.2 Making data openly accessible</p>	<ul style="list-style-type: none"> • Most of the scientific data will be openly available, at least those having been published in international scientific journals or announced in international conferences and published to their minutes. Some of the sensitive administrative data cannot be available, i.e. the financial data of the budget of the participating institutions. This restriction is also implemented in view of internal rules and regulations hold by each institution and current European law concerning these matters. • The data will be made available in a repository created in the web. • The usual methods or data base software tools (e.g. in open source code) in the market needed to access the data for documents and designs will be used. A proper documentation about the software needed to access the data will be included. • A repository owned by the CompactLight collaboration will host all kinds of open data and associated metadata, documentation, and code created by CompactLight for dissemination. • There will be well-described guiding for access to the open data without restrictions.
<p>2.3. Making data interoperable</p>	<ul style="list-style-type: none"> • The open data of the XLS project will be interoperable, allowing data exchange and re-use between the researchers, institutions, organizations, countries, etc. For this reason, there will be adhered standard formats as much as possible compliant with available open software applications and in particular facilitating re-combinations with different datasets from different projects. • The terminology of the scientific data and metadata will be used following standard vocabulary to all data types and methodology focusing to have our data interoperable.
<p>2.4. Increase data re-use (through clarifying licences)</p>	<ul style="list-style-type: none"> • As far as possible there will be no licence to our data access, in order to permit the widest possible re-use. • Most of the data of the CompactLight project will be available immediately after approval by the researchers of the project collaboration. An effort will be made to reduce the time for publication or patent request as much as possible, given that the publication and patent procedure in Europe are quite fast. • The data will be used or useable by third parties and in particular after the end of the project without restriction for re-use. • There will be described data quality assurance processes for the CompactLight data. • The resources for long term preservation will be discussed and approved by the XLS collaboration, so the data will be available for a long duration.

<p>3. Allocation of Resources</p>	<ul style="list-style-type: none"> • There is yet no precise estimation of the costs for making the data of the project FAIR. Costs will be estimated taking into account a data base server with the maintenance plus one FTE for the total period of the project duration. • The responsible for the data management will be the XLS scientific coordinator. • The resources for long-term preservation will be discussed and approved by the XLS collaboration.
<p>4. Data Security</p>	<ul style="list-style-type: none"> • The data security is split into two concepts: data security for the storage media with multi-type backup repositories and data security for people that access and read data without being able to harm them.
<p>5. Ethical Aspects</p>	<ul style="list-style-type: none"> • Currently no ethical or legal issues related to open data sharing are expected.
<p>6. Other</p>	<ul style="list-style-type: none"> • <u>No other national/funder/sectorial/departmental procedures for data management will be used.</u>

<p style="text-align: center;">HISTORY OF CHANGES</p>		
Version	Publication date	Change
1.0	29.06.2018	<ul style="list-style-type: none"> ▪ XLS: Data Management Plan v1.0 – 29.06.2018