



Funded by the
European Union

Compact 

XLS-Report-2018-001
21 December 2019

XLS Deliverable D1.2

CompactLight Data Management Plan V.1.2

G. D'Auria^{*}, E. Gazis^{1)§}, A. Latina[†], R. Rochow^{*}, M.
Aicheler[‡]

On behalf of the CompactLight Partnership

Prepared on: 21.12.2019

^{*} Elettra, Italy, [†] CERN, Switzerland, [‡] UH/HIP, Finland, [§] IASA, Greece

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.

¹Corresponding author: evangelos.gazis@cern.ch



Funded by the
European Union

Compact 

© 2019 Elettra Sincrotrone Trieste for the benefit of the CompactLight Collaboration.
Reproduction of this document or parts of it is allowed as specified in the CC-BY-4.0 license.

Information

Project Number: 777431
H2020-INFRADEV-2016-2017/H2020-INFRADEV-2017-1

Project Acronym: **XLS**
Project title: CompactLight

Contents

- 1 Introduction** **6**
 - 1.1 FAIR Data Management 6
 - 1.2 Structure of the DMP 6

- 2 Data Summary** **7**

- 3 FAIR Data** **8**
 - 3.1 Making data findable, including provisions for metadata 8
 - 3.2 Making data openly accessible 8
 - 3.3 Making data interoperable 9
 - 3.4 Increase data re-use (through clarifying licences) 9

- 4 Allocation of Resources** **10**

- 5 Data Security** **11**

- 6 Ethical Aspects** **12**

- 7 Other Issues** **13**

History of Changes		
Version	Publication Date	Change
1.0	29.06.2018	XLS: Data Management Plan v1.0, 29.06.2018
1.1	05.04.2019	Inserted in the XLS LaTeX Template
1.2	21.12.2019	Overall Update and implementation of EDMS specific information

1 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.

1.1 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.

1.2 Structure of the DMP

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

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.

2 Data Summary

The purpose of the data production, collection and processing is to gather all the scientific information produced by our XLS collaboration, which is defined as “open” by the partners. We are collecting scientific as well as administrative data in order to deliver a coherent and fully documented Conceptual Design Report (CDR) at the end of the project. All this data could be used subsequently by other researchers and hence 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 is cast into our working data formats, e.g. we will reuse previous designs and measurement data as references to our design and development processes. Most of the work the collaborators are conducting is in their field of previous expertise and work. Hence, most of the data comes from the scientific and engineering work carried out previously and currently 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. The total volume of centrally administered and archived data is estimated lower than 1 TB. 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 and alternative way of electron acceleration, developed from the CLIC technology. The ultimate deliverable, the CDR of our project and its related data, will offer novel and innovative XFEL designs with compact, low-sized accelerators as well as low construction and operation cost.

3 FAIR Data

3.1 Making data findable, including provisions for metadata

CompactLight is using CERN's own Engineering & Equipment Data Management Service (EDMS) which CERN uses as its main Product Lifecycle Management (PLM) system and Enterprise Asset Management system. Much more details on how CERN is using this system, features and tutorials can be found under <https://edms-service.web.cern.ch/>.

EDMS enforces unique identifying numbers. The system allows for direct search of the respective EDMS numbers as well as any query containing author, keyword, creation date range. Additionally, documents in EDMS are attached to (multiple) hierarchical projects. Individual documents can hence be found also through simple navigation of the relevant projects and nodes.

For official deliverables, documentation and publications, we are following the naming convention of e.g. "XLS-D1.2-DMP". Documents will hence be identified with project name, deliverable number and document name. For all science related documents, we do not follow a particular naming convention.

The project does not intend to use specific search keywords for optimizing possibilities for re-use of our data. Modern information technology archive and query systems allow for such vast and flexible searches, that the model of "keywords" is considered being outdated.

The project data storage and archiving system EDMS has an in-build function of versioning, visually displaying all previous states of documents and allowing only the current one to be altered. For documents for which we specifically have multiple updated versions planned, e.g. DMP, we explicitly add version numbers in the documents themselves. The project is documenting any information available and necessary in order to be able to recreate the entire design process. One example is simulation output data which is complemented with simulation code documentation as well as input parameters and tools/functions used. Specifically, for the numerical simulations for which tailored computer codes are being used we use CERN's GitLab environment with all its Git features and possibilities. For all more standard data sets, e.g. engineering design and cost analysis, EDMS allows for an exhaustive set of predefined information fields where useful meta data information can/must be attributed to the respective documents directly.

3.2 Making data openly accessible

Most of the scientific and technical data will be openly available once finalized, such as technical specifications or simulation results. Snapshots of various developments and data sets will be published in international scientific journals or presented in international conferences and published in their proceedings. Some of the 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.

The open data will be accessible through the CERN hosted EDMS system (as described above). The web based system allows for direct queries or browsing through hierarchical structures in order to find the desired data set or information. No specific software tools are necessary and access to the service is available from all over the world. There is an extensive guide on how to use EDMS readily available at CERN under <https://edms-service.web.cern.ch/edms-service/faq/EDMS/pages/tutorials.html>. The CERN hosted EDMS system is free to use for CERN related activities (XLS qualifies for this criteria through the role of being a partner of the project) and the XLS data will be saved, archived and backed up according to CERN standard procedures along with all the rest of the CERN engineering data.

There will be no restrictions for access to open data. As for the sensitive data, access can be granted through the CERN implementation of e-groups. Potential users of the data can obtain access rights through acquiring a CERN light weight computer account and being added to the relevant e-group. For the duration of the project, the access for externals will be granted by the project coordinator on a on-demand case.

3.3 Making data interoperable

Most of the scientific data and administrative data will be cast in the form of reports or publications, hence usable with any office productivity suite. Raw data of measurements (where applicable) or simulation output will be stored in plain text, ascii file format or similar, containing headers for data type identification. Hence special software will be necessary for performing post-project analysis. If requested, specifically developed computer codes can be provided for cross checking of results.

The project uses mainly vocabulary and expressions common in the light source as well as accelerator communities where XLS as a project is situated and its partners are coming from.

3.4 Increase data re-use (through clarifying licences)

The scientific data produced by CompactLight will be accessible without licence immediately after publication. Some of the more sensitive administrative data (such as company offers, cost and schedule estimations, etc.) will be freely accessible at latest after the end of the project. Access to and the maintenance of the full data set will be ensured according to the data policy of CERN.

4 Allocation of Resources

The use of EDMS within the context of CompactLight is free of any cost for the project. Work package 1 is responsible for the development and the implementation of the data management plan and hence the man power is covered by the partners of the respective work package. As mentioned above, long term preservation will be provided according to CERN methods and practices.

5 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 implemented and handled by CERN's EDMS system, while data security with respect to data handling by users will be ensured through selected access rights for limited to reading and downloading data sets without the possibility of changing / manipulating / saving them in the repository.

6 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 expressively agreed with the concerned persons.

7 Other Issues

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

Link to instructions and template:

https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf