Ref. Ares(2021)5934997 - 29/09/2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964588.

Data Management Plan

Deliverable number:	D4.2
Due date:	30/09/2021
Nature[1]:	ORDP
Dissemination level[2]:	PU
Work Package:	WP4
Lead Beneficiary:	POLIMI
Contributing beneficiaries:	CNR, INCDTIM, C5

[1] R = Report, P = Prototype, D = Demonstrator, ORDP = Open Research Data Pilot, O = Other, W = Websites, patents filling, etc.

[2] CO = Confidential, only for members of the consortium (including the Commission Services)

PP = Restricted to other programme participants (including the Commission Services)

PU = Public

RE = Restricted to a group specified by the consortium (including the Commission Services)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964588.

Version	Date	Description
Version 1	21/09/2021	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964588.

Summary

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964588.

1. Preamble

This Data Management Plan (DMP) was drawn up following the Horizon 2020 FAIR Data Management Plan (DMP) Template.

The DMP will allow X-PIC partners to address all issues related to data. However, this document is intended to be a living document that will be updated as the implementation of the project progresses and when significant changes occur.

The following sections provide detailed information on the data sets, standards and metadata, and the respective data sharing and archiving and preservation procedures for the data sets collected at each partner institution.

2.1 Data Summary

This project aims at the realization of a new technological platform for the development of a EUV – soft X-ray integrated photonics.

Different types of data and formats, depending on their specific origin, will be generated and collected, as summarized by the following categories:

(i) images representing magnified photos of the X-PIC micromachined devices; fluorescence pictures of laser-induced plasmas produced inside the devices; results of photolithographic processes on semiconductor samples.

(ii) 1-D arrays representing spectra of the laser source; temporal fluctuation of the laser parameters; temporal evolution of lab environmental disturbances.

(iii) 2-D maps representing spectra of the XUV-soft X radiation generated inside the devices and resolved in wavelength/photon energy and divergence; slices of the computed speed, pressure, temperature, or density of the gas flowing inside the devices; slices of the simulated generation efficiency of the radiation inside the devices.

(iv) 3-D maps representing hyperspectral images of semiconductor samples probed by the X-PIC devices.

(v) Text files representing software developed for remote control of the experimental setup, for data analysis and numerical simulations.

Images will be in .jpg, .png or .tiff format; extension of software files will correspond to the specific programming language (.f, .m, .c, .cpp, .py etc.). 1-D, 2-D and 3-D arrays will be stored in hierarchical data format (.h5) allowing the storage of metadata inside the same file.

The origin of data will mostly rely on spectrometers, CCD cameras, sensors, numerical simulations.

The main aim of the data collection/generation is to design, test and demonstrate applications of micromachined photonics devices operating in the XUV-soft X.

The expected size of the data may vary from a few kB per file to GB per file according to the file nature.

The data may be useful to the scientific community for comparing experimental and theoretical outcomes and for benchmarking results achieved in different laboratories.

The following tables describe the specific data generation and use for each partner.

POLIMI- Partner 1	
What is the purpose of the data collection/generation and its relation to the objectives of the project?	Spatial and spectral properties of the HHG emission from the devices, temporal characterization of the emitted pulses, measurement of the generation efficiency
What types and formats of data will the project generate/collect?	.m, .c, .cpp, .py, .h5 formats
Will you re-use any existing data and how?	No
What is the origin of the data?	CCD camera, XUV/soft X spectrometer, VMI electron spectrometer, IR/visible spectrometer, oscilloscope, FROG device

What is the expected size of the data?	From 1 kB to 1 GB
To whom might it be useful ('data utility')?	All the partners working on the X-PIC project; scientific community working on HHG and related applications

CNR- Partner 2	
What is the purpose of the data collection/generation and its relation to the objectives of the project?	Characterization of the fabricated devices; Numerical simulations concerning the hollow waveguide guiding properties. Laser source characterization; HHG spectra;
What types and formats of data will the project generate/collect?	Tiff, jpg, txt, h5 Common image files formats (e.g. bmp, jpeg, tif, pdf, dvi), common files for lithography (e.g. gds, dxf, cvs)
Will you re-use any existing data and how?	No
What is the origin of the data?	Microscope images, near field and far field of coupled modes, simulation of hollow waveguide guiding properties. Instruments for the characterization of laser sources, such as photodiodes, spectrometers, oscilloscopes. XUV spectrometers
What is the expected size of the data?	From 1kB to 10 GB
To whom might it be useful ('data utility')?	Experimental teams working on: i) femtosecond laser micromachining, ii) integrated optics, iii) HHG, iv) nano-micro and soft lithography.

INCDTIM partner3

What is the purpose of the data collection/generation and its relation to the objectives of the project?	Numerical simulations concerning pulse propagation, harmonic generation and XUV propagation in hollow core waveguides
What types and formats of data will the project generate/collect?	Text files or binary files containing numerical data, pdf files and image files containg results.
Will you re-use any existing data and how?	No, the data are specific to the X-PIC project
What is the origin of the data?	Codes written by us.
What is the expected size of the data?	From 1 kB to 10 GB
To whom might it be useful ('data utility')?	Experimental teams working on: i) laser design, ii) fiber fabrication and iii) XUV generation

C5- Partner 4	
What is the purpose of the data collection/generation and its relation to the objectives of the project?	Purpose: Developing, monitoring and testing of OPCPA system and its components. Relation: Guarantee required specifications of the light source and determine parameters for subsequent experiments.
What types and formats of data will the project generate/collect?	Typically, spectra, beam profiles, autocorrelation traces and power stability data will be generated including, .wcf, .txt, .jpg, .png and similar file formats.
Will you re-use any existing data and how?	Existing data could be re-used for crosschecking the performance of the developed system with previous OPCPA systems.

What is the origin of the data?	The data originates from spectrometers, cameras, power meters, photo diodes or retrieval algorithms in case of pulse duration measurements.
What is the expected size of the data?	Approximately several Gb.
To whom might it be useful ('data utility')?	The consortium partners.

2.2 FAIR data

2.2.1 Making data findable, including provisions for metadata

The data produced within X-PIC will be made available in a public repository which provides a Digital Object Identifier (DOI) for each dataset, making the data identifiable and locatable via a persistent and unique identification mechanism. Thus, we assure datasets to be identifiable and citable in accordance with the FAIR principles. Such DOI will be used to cite data related to the scientific publication outcomes, enhancing reproducibility and dissemination of our results.

We will follow standard naming conventions, that will be defined accordingly to the project development. The DMP will be updated with the official nomenclature-definition when the testing phase has been standardized.

Keywords will be always carefully defined per each dataset submitted, so that other users could easily find and re-use our datasets. To this aim, metadata will be linked to each dataset in the form of digital documentation; hierarchical data format will be used whenever possible in order to include data and metadata in the same file.

2.2.2 *Making data openly accessible*

The data related to publications will be made available once the demonstration of the project related results has reached publication in peer-reviewed scientific journals, and in compliance

with the embargo policy of the journal (but, in any case, no later than 6 months after publication, in agreement with H2020 open access policy).

Other data may be made available provided that the disclosure does not prevent subsequent exploitation of the project results, including patents.

The data will be accompanied by a documentation file describing the data content and the experimental/theoretical conditions upon which the data have been generated.

In view of the formats used for our data, no proprietary tool is needed to access the data.

The data will be made accessible by loading in a public database service such as Zenodo and linked to OpenAIRE (this will ensure the possibility to assign a DOI to each of them).

2.2.3 Making data interoperable

The data produced and publicly released will be interoperable by default, adhering to standard formats for the image and code data previously described. Standard vocabulary for all data types will be used. This will speed up novel developments and applications in the interested communities.

2.2.4 Increase data re-use (through clarifying licences)

The Grant Agreement provides for three communication impact exploitation and dissemination activities plans to ensure the usability of the data produced by the project and one market research report for future impact exploitation and dissemination strategy.

Data made available is also re-usable if data origin is fully acknowledged. Restrictions apply to use by third parties of data produced in the project when the disclosure prevents subsequent exploitation of the project results; such data will not be available on open repositories. Data published on open repositories will be permanently re-usable.

3. Allocation of resources

We foresee minimal costs for making data FAIR. We will mostly rely on freely available tools for loading and storing the project outcomes.

The person in charge of the data management will be Prof. Salvatore Stagira.

We will monitor the data availability during the entire duration of the project and a further period of time of two years after the ending of the action. Ideally, any published data would be kept accessible for an indefinite amount of time and in the unlucky event of repositories' policy-change, the authors would provide direct support to those requesting any data previously available.

4. Data security

All the data will be stored in local online servers equipped with redundant arrays of independent disks and accessible through a virtual private network to prevent data loss related to malicious access. Where possible, additional redundant storage will be performed on a crypted portable external hard-disk to further minimize data loss events.

Publishable data will be also stored in the online Zenodo repository. We will make use of the available institutional (automated) backup services provided by the IT service at the collaborating organisations or at least keep several backups, to maximally reduce the data loss. In the case the data previously published online will be lost for any reason, we will provide to upload it again to another repository such as figshare (https://figshare.com/), dryad (https://datadryad.org/) or OSF (https://osf.io/).

Dropbox will be used for sharing of not sensitive data among project partners. Confidential data, whose disclosure to third parties might hinder the outcomes' exploitation, will be shared in more secure ways like cryptographic public/private keys systems.

5. Other issues

Part of the scientific research results will be made available on the project site https https://www.x-pic.eu/ to ensure the dissemination and on the institutional repository of POLIMI https://re.public.polimi.it/, of INCDITM https://www.itim-cj.ro/xpic/ and of C5 https://www.class5photonics.com. CNR will use ZENODO.