



# APPROACH

Advanced Photonic PRocesses for novel sOlar energy hArvesting  
teCHnologies

## Deliverable

---

### D1.2 Data management plan

Deliverable Lead: FORTH

Deliverable due date: 30 November 2023

Actual submission date: 13 December 2023

Version: 1.0



This project receives funding in the European Commission's Horizon Europe Research Programme under Grant Agreement Number 101120397

<b>Document Control Page</b>	
<b>Title</b>	Data management plan
<b>Creator</b>	FORTH
<b>Description</b>	1st Data management plan draft
<b>Creation date</b>	11 December 2023
<b>Type</b>	Report
<b>Language</b>	English
<b>Audience</b>	<input checked="" type="checkbox"/> public <input type="checkbox"/> confidential
<b>Review status</b>	<input type="checkbox"/> Draft <input type="checkbox"/> WP leader accepted <input checked="" type="checkbox"/> Coordinator accepted
<b>Action requested</b>	<input type="checkbox"/> to be revised by Partners <input type="checkbox"/> for approval by the WP leader <input type="checkbox"/> for approval by the Project Coordinator <input checked="" type="checkbox"/> for approval by the Project Officer

## Table of contents

Abstract.....	5
1. Introduction.....	5
2. Overview of the Data Management Plan.....	6
3. FAIR Data.....	7
3. Datasets.....	9
3.1 Data summary.....	9
3.2 Datasets Forms.....	9
3.3 Storage and access.....	14
4. Ethical and Legal issues.....	15
4.1 Ethical Issues.....	15
4.2 Security and Confidentiality.....	15
4.3 Ownership-Intellectual Property Rights (IPR).....	15
4.4 Protection of personal/special categories of data.....	16
5. Dissemination.....	16
Conclusions-Implications.....	17

## DISCLAIMER

The sole responsibility for the content of this document lies with the APPROACH project and in no way reflects the views of the European Union.

## Abstract

The purpose of this document is to present the preliminary Data Management Plan (DMP) for the APPROACH project, outlining the data management policies that will be employed by the project partners to handle the datasets produced throughout it. The DMP identifies the primary research data to be generated during the project's execution and outlines the actions of collecting, processing, curating, and preserving the data throughout and after the project's conclusion. It also outlines how the data will be shared with both the scientific community and APPROACH partners, as well as how it will be made accessible for verification and reuse. The DMP's objective is to promote efficient data management among partners and ensure that the data remains FAIR (Findable, Accessible, Interoperable, and Reusable) for future researchers and scientific communities. The DMP is a dynamic document that will be updated and improved as the project progresses. It is worth noting that the DMP aligns with the project contracts and serves as a supplement to the proposed exploitation, dissemination, and IPR procedures.

## 1. Introduction

Partners have considered all potential issues related to management of both open and commercial data as they want to maximise data security, accessibility and validation of results whenever possible. Besides, they want to ensure that the funding and regulatory body requirements are met and the datasets remain accurate, authentic, reliable and complete. A Data Management Plan (DMP) has been developed covering the entire data management life cycle, in line with the “Guidelines on Data Management in Horizon Europe”, for making all produced data Findable, Accessible, Interoperable and Reusable (FAIR) while it will minimise possible duplication of effort and the risk of data loss. The processing of personal data will comply with the legal obligations set at the EU level by the General Data Protection Regulation (GDPR) and at the level of the single Member States by the national laws on data protection. Guided by ethical considerations and the EU values and principles on which the EU is founded, this project will ensure data quality, safety, and confidentiality. Data generated will include among other measurements, protocols, specifications, software and mathematical models. These data include datasets identification and description, standards and metadata, data exploitation/sharing for verification and re-use and data archiving/preservation. Moreover, the types of data/research outputs that will be generated will include newly collected experimental data with metadata descriptions (text, image, numerical, video), analysis files such as through OriginLab or Graphpad Prism, as well as protocol documents, lab notes, reports, presentations, protocols, specifications, software and mathematical models. Where possible (nearly all cases), we will use standard, open formats (.txt, .ascii,

.csv, etc) to ensure interoperability. The data management plan describes procedures for creating, organising, documenting, storing and sharing various data produced. It takes into account issues such as data protection and confidentiality, data preservation and curation and provides a framework that supports researchers and their data throughout the course of their research and beyond. The DMP defines and implements the privacy and data protection policies. The DMP has been formulated, in compliance with the FAIR guidelines in all aspects of scientific data management and according to the exploitation strategy and the IPR requirements as has been defined in the Consortium agreement. A set of recommendation and procedures is available for all users with the aim to have as many as possible FAIR datasets published as open data.

This document contains guidelines that will be used for the development of a Data Management Plan (DMP) which will include an analysis of the main elements of the data management policy that will be used by the consortium with regards to all the data that will be generated by the project.

Moreover, the DMP covers the following aspects:

- Description of the data to be collected / created
- Standards / methodologies for data collection and management
- Ethics and Intellectual Property concerns or restrictions
- Plans for data sharing and access
- Strategy for long-term preservation

The DMP will not be a fixed document, but it will evolve and will gain more precision and substance during project implementation. New versions of the DMP will be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors. More detailed versions of the DMP will be delivered at later stages of the project.

## 2. Overview of the Data Management Plan

### *Purpose of the Data Management Plan*

The Data Management Plan (DMP) purposes are:

- to support the data management life cycle for all data that will be collected, processed or generated by the Project,
- to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the Project,

- to provide detail and guarantee about the preservation of the data collected during the Project, as well as any results derived from the associated research,
- to provide detail on how we plan to address the Ethical issues (if any) related to data that will be collected during the Project timeframe,
- to create a document which explains the management of data collected during the Project.

The DMP is not a fixed document, but it will evolve during the Project.

APPROACH is a coordination and support action, therefore project data will be collected and generated as a result of achieving the main objectives that are:

- 1) Increase research excellence of the Widening partners in the field of research under ERA-Talents workframe, as short-term exchanges, expert visits, joint summer schools, seminars and workshops to increase mobility of qualified scientists, upgrade competences and raise the experience level of existing personnel;
- 2) Enhance the reputation, attractiveness and networking channels of the Widening partners by participation and organisation of international conferences, seminars, workshops, networking, and other events, as well as increase percentage of scientific articles published;
- 3) Enhance the scientific and technological capacity of the linked institutions, focused on the Widening Countries, establishing data management platforms for multidisciplinary research.

### *Intended audience*

The DMP will be oriented to:

- the project's participant organizations;
- the local Ethics Committee;
- the partners' personnel and all stakeholder involved by the Project;
- the European Commission.

## 3. FAIR Data

Specific measures will be applied during the evolution of the APPROACH project in order to ensure the proper application of the FAIR principles, in line with the relevant EC guidelines, i.e., ensure data that is findable, accessible, interoperable and reusable:

### *Making data findable, including provisions for metadata*

In order to make data findable, metadata will be used. All partners have agreed in providing relevant metadata and keywords, so that their data will be easily discoverable. Clear version numbers will be

included (automated process through the project repository) and standard naming conventions will be defined.

### *Making data openly accessible*

The consortium partners already identified which data will be made openly available and which cannot be shared (or needs to be shared under restriction), including the reason why access is restricted in the latter case. In the next DMP versions, more details will be provided regarding the accessibility of the data. Details on the repository, the methods and tools necessary to access the data will be included in future versions of the DMP. The consortium aims at using the project website (or other easily accessible repositories) as repository for open accessible (public) data, ensuring easy access to anyone interested.

### *Making data interoperable*

Provisions are also taken to make data interoperable, making it easier to exchange and re-use them across research institutions, organisations, etc. The project strives in making all open data interoperable. Data will be made available in standard formats (e.g. .doc, .xls, .jpg, .tiff, .pdf) compatible with available (open) software applications in order to enable unrestricted data exchange between researchers, institutions, organisations, countries and others. Inter-disciplinary interoperability of data will be enabled by standardized and organized controlled vocabularies for metadata description.

### *Increase data re-use*

The public content made available via the project website will be available for download and re-use with no restrictions or embargo. Public data will be made available for re-use. The consortium partners will attach specific licenses to the deposited data to define all conditions under which the work is provided, under either open or restricted access. The Widening actions data (e.g., transcripts of workshops and training activities, surveys, questionnaires and experts' interviews) will remain reusable by delivering them in standardized, wide-spread data formats (mp3 and pdf/a). There is no embargo foreseen. The data will be released with a clear and accessible open data usage license which will allow the further processing under attribution and sharing under the same conditions.

### *Allocation of resources and responsibilities*

Presently, no specific additional costs are foreseen beyond the described processes and methods for the data collection and storage set out in this DMP and which fall within the activities covered by the current grant. Later versions of this DMP may identify additional costs if required and properly justified by the project's future activities and needs. The project coordinator (FORTH) is in charge of the DMP from both the scientific and technical perspectives, as well as the release of the first version and the regular updates. The partner responsible for the generation of data in each WP will be also in charge of the validation and registration of datasets and metadata, as well as backing up data for sharing through open access repositories, while quality control of these data is the responsibility of the relevant

WP leader, supported by the Project Coordinator. Each partner should respect the policies set out in this DMP.

### 3. Datasets

#### 3.1 Data summary

The training and transfer of cutting-edge experimental and theoretical techniques will be done using already existing scientific data, in the process of exchange of scientific knowledge and data. However, various type of data will be collected and/or created in the context of the project activities, such as:

- Data on project management and coordination: data about the project partners, data generated by the consortium meetings, reporting data related to project management. Key performance and output indicators for assessment of the APPROACH scientific improvement and impacts.
- Data on project's Widening actions: pool of science related material that originates from the project's Widening actions, pool of material that originates from the meetings with policy makers and industrial stakeholders, reporting data generated by the Widening actions of the project.
- Data on workshops and summer schools for early-stage researchers - pool of science related material and reporting data that originates from the project's networking and training activities.

Moreover, the Data Sets for the new generated knowledge derived from research collaborative activities will be identified and described with the contribution of all the project partners. A short description of the data that will be generated in the research project (e.g., samples, physical collections, software, materials, to be produced during the project) must be provided. Additionally, an estimation of the amount of data and content of the data (if possible) must be included.

In order to collect all relevant information on the project's data mentioned above, a dedicated table was generated for each dataset and provided to the consortium partners. Topics covered in these tables include all relevant information, such as description of identification, responsible partners, use of metadata, definition of data formats, provisions to make data FAIR, security and ethical issues. Representative tables with detailed information on each data set are provided below.

#### 3.2 Datasets Forms

##### Dataset Type 1 - Project management and coordination data

PROJECT MANAGEMENT AND COORDINATION DATA	
Data identification	Data description

Data about the project partners	Mailing list, address and bank details for the distribution of the financial contribution of the Financing Authority
Consortium Meetings' Data	Data generated by the meetings
Reporting data	Data on using/reallocation of the budgetary resources, data on the progress of the project activities according to the respective tasks, milestones and deliverables, data on the actual performance against the project's objectives and KPIs
Source of data	Data provided by the project partners
<b>Partners responsibilities</b>	
Partner in charge of data collection	FORTH
Partner in charge of data analysis	FORTH
Partner in charge of data storage	FORTH
<b>Standards</b>	
Info about metadata (production and storage dates, places) and documentation	All relevant reporting data and documents will be stored, as appropriate, in dedicated project files, or electronically, through the reporting interfaces in the project's drive folder.
Formats	Various forms, e.g., paper, photos, videos, electronic documents.
<b>Data exploitation and sharing</b>	
Data exploitation (purpose/use of the data analysis)	The data will be used to monitor the progress of the project implementation and prepare project progress reports, as well as for financial control and payments.
Data access policy / Dissemination level	All data collected from the project partners will be kept confidential. The official periodic reports (technical and financial) and project deliverables submitted to the EC are confidential.
Data sharing, re-use, distribution, publication	The reporting data will be shared by the project partners electronically through the project's drive folder. The official periodic reports and the minutes of the consortium meeting will be submitted through the EC Funding & Tenders Portal.
Embargo periods (if any)	N/A
Personal data protection	Personal data of the project team personnel, Supervisory Board, Coordination Committee, (mailing list, etc.) are for internal use only.
<b>Archiving and preservation (including storage and backup)</b>	
Data storage (including backup)	The data sets used in the project management and coordination will be stored during the project implementation, as appropriate, in dedicated project files, or electronically in the project's drive folder. The relevant datasets relating to these activities will be stored in dedicated databases for at least 3 years period after the end of the project.

**Dataset Type 2 – Widening Actions (including workshops and training activities)**

Widening Actions	
Data identification	Data description
Pool of science related material that originates from the project’s Widening actions	Videobooks/pictures/ of transferable trainings including lectures and presentations from the project workshops on scientific and technological capacity building
	Video/picture files from webinars resulted from the short-term staff exchanges and visits
	Video/pictures files from webinars resulted from the short-term expert visits
	Material from the Scientific knowledge exchange
Pool of material that originates from the meetings with stakeholders	Material from the “Bridging the gap” between research and innovation workshops
Reporting data generated by the networking and training activities of the project	Data on the project workshops’ results, lists of lecturers and participants, workshops’ agendas, data on the short staff exchanges and short terms expert visits, etc.
Source of data	Data provided by the responsible Task Leader
Partners responsibilities	
Partner in charge of data collection	Responsible Task Leaders and FORTH (the coordinator)
Partner in charge of data analysis	All Partners
Partner in charge of data storage	FORTH
Standards	
Info about metadata (production and storage dates, places) and documentation	For the purpose of the project documentation this data will be stored individually by each responsible partner.
Formats	MS Office format (docx, xlsx, pptx) and pdf, images (svg, png, jpg), videos (mp4)
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	These data will be used for developing the related public deliverables of the project and for communication and dissemination purpose.
Data access policy / Dissemination level	The reporting data provided by the project partners will be kept confidential. The protection and procedures about the use of IPRs will be applicable to outputs explicitly resulting from the project activities. The reports and other related deliverables on the Widening actions will be public but will only include non-sensitive data.

Data sharing, re-use, distribution, publication	The reporting data will be shared by the project partners electronically through the project drive folder. The project deliverables will be submitted to the EC Funding & Tenders Portal. The public deliverables of the project will be published on the project website. The material for communication and dissemination resulted from the projects' networking and training activities will be published also on the project website.
Embargo periods (if any)	N/A
Personal data protection	Data collected on the participants attending workshops, meetings and training activities will be limited to required registration data, i.e. name, name of organisation, faculty or department, role in the organisation, and contact data (email address). An Informed Consent Form explaining the use of the personal data will be available for consultation as part of the registration form and its approval will be necessary to complete the subscription.
<b>Archiving and preservation (including storage and backup)</b>	
Data storage (including backup)	The data sets will be stored during the project implementation, as appropriate, in dedicated project files, or electronically in the project's drive folder. The relevant datasets relating to these activities will be stored in dedicated databases for at least 3 years period after the end of the project.

### Dataset Type 3 – Research and experimental data

Main dataset (name)	Optical_spectroscopy
Sub-dataset (name)	PLspectroscopy
Responsible partner	FORTH
Other partners involved	UPOL, IMEC
Goal	Collect, save and secure all relevant information extracted from experimental runs of scientific projects related to APPROACH
Data origin-purpose	Monochromator software. Assess optical quality via collection of the photoluminescence emission of the materials in the energy, wavelength, time domains.
Standards and metadata	Type of optical spectroscopy, experimental project name, material, temperature, date
Data collection	Originlab
Dataset type	SPE, opj
File formats	.txt, .ascii, .csv
Expected volume of data	1Gb
Expected time of release	M24
Keywords	PL, TMD

Data quality	High, Data control will be performed by replication of samples and measurements following standardized data acquisition, and peer review of data for their publication.
--------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Main dataset (name)	Experimental techniques for materials characterization
Sub-dataset (name)	Materials characterization data
Responsible partner	UPOL
Other partners involved	FORTH, IMEC
Goal	Collection of data derived from experimental techniques for materials characterization
Data origin-purpose	Experimental techniques for materials characterization (such as surface analysis - XPS, ambient and low temperature photoluminescence spectroscopy, light-induced electron paramagnetic resonance, NMR and gas/liquid chromatography for product analysis)
Standards and metadata	Type of Material, Date, Number of experiments, type of characterization
Data collection	Raw data will be collected through specific softwares associated to the instruments and exported to non-proprietary formats. Then they are converted to *.tif, *.jpeg, *.bmp, *.png. (EDS) files which are usually converted to *.html, *.jpg files, *.pdf or even *.txt files.
Dataset type	Experimental data
File formats	*.dat, *.txt, *.tiff, *.jpeg, *.bmp, *.png, *.pdf
Expected volume of data	2 Tb
Expected time of release	M30
Keywords	Spectroscopy
Data quality	High, Data control will be performed by replication of samples and measurements following standardized data acquisition, and peer review of data for their publication.

Main dataset (name)	Microscopy Techniques for advanced materials
Sub-dataset (name)	Microscopy images
Responsible partner	UPOL
Other partners involved	All
Goal	Assessment of nano-, micro- crystal average size, build-up of a size distribution histogram (for low-magnification images); confirmation of crystal phase (for high-magnification, phase contrast images); compositional mapping (for EDS maps).
Data origin-purpose	High-resolution transmission electron microscope, atomic force microscope and scanning electron microscope results. Generated data within the project will allow to assess the size and morphology of the synthesized nano- and micro- crystals, as well, as their chemical composition.
Standards and metadata	Type of Material, Date, Number of experiments

Data collection	Raw data will be collected through specific softwares associated to the instruments and exported to non-proprietary formats. For example, TEM images will be collected and exported preferentially as *.dm3 or *.tiff uncompressed files. Then they are converted to *.tif, *.jpeg, *.bmp, *.png. (EDS) files which are usually converted to *.html, *.jpg files, *.pdf or even *.txt files.
Dataset type	Experimental data
File formats	*.jpeg, *.tif, *.bmp, *.png, *.pdf, *.txt
Expected volume of data	100 Gb
Expected time of release	M36
Keywords	TEM, HRTEM, EDS, SEM, HR-SEM, EDS mapping
Data quality	Data control will be performed by acquiring images from different replicated samples, at different magnification, from different regions, and following standardized acquisition conditions to avoid beam damage on dried samples

### 3.3 Storage and access

For the duration of the project, datasets will be stored on the responsible partner's centrally provided storage. To ensure the safety of the data, the involved participants will use their available local file servers to periodically create backups of the relevant materials. Additionally, all other relevant documentation created during the project such as deliverables will be self - archive and preserved in APPROACH Sharepoint Drive folder of FORTH that has been created for the purposes of the project. It allows users to store files in the cloud, share files, and edit documents, spreadsheets, and presentations with collaborators.

The APPROACH Sharepoint Drive Folder at FORTH will be accessible by all of the partners of the APPROACH consortium and it will include the following subfolders/architecture:

- Administration (where all the files related to administration issues will be stored like the Grant Agreement, the Consortium Agreement, administrative templates, information related to meetings etc).
- Work Package Documents (containing one folder for each WP in which data relevant to each WP will be included including all the deliverables- drafts in the beginning that will be substituted by the finals when submitted- and all other relevant data, files and folders)

All of the research data and material will be in place for at least the 2 years after the end of the project prescribed by the European Commission, as well as the foreseeable future following that according to the agreements reached by the consortium by the end of the project (if any additional is agreed). After completion of the project, all the responsibilities concerning data recovery and secure storage will go to the repository storing the datasets. Data selected by the partners for sharing with the public will be

archived and preserved in the Zenodo data sharing repository. The Coordinator of the APPROACH project along with the Dissemination & Exploitation Manager is in charge for data management and all the relevant issues.

## 4. Ethical and Legal issues

### 4.1 Ethical Issues

The APPROACH partners are to comply with the ethical principles as set out in Article 14 of the Grant Agreement, which, among other, states that all activities must be carried out in compliance with:

- (a) ethical principles (including the highest standards of research integrity – as set out, for instance, in the European Code of Conduct for Research Integrity – and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct) and
- (b) applicable international, EU and national law.

### 4.2 Security and Confidentiality

All data will be stored and transferred according to applicable national, EU and international legislation for data security regulations. General procedures for data handling, management and storage will be applied.

All APPROACH partners must keep any data, documents or other material confidential during the implementation for the project and for four years after end of the project in accordance with Articles 13, 15, 16 of the Grant Agreement. Further detail on confidentiality can be found in these Articles of the Grant Agreement.

### 4.3 Ownership-Intellectual Property Rights (IPR)

In the DMP issues regarding copyright and Intellectual Property Rights of the data are included. These issues are set in the Consortium Agreement and the Grant Agreement of the APPROACH project regarding all the results of the project. Thus, the DMP follows the Consortium Agreement and the Grant Agreement that is signed by all project partners regarding Ownership issues. Materials generated under the APPROACH Project will be disseminated in accordance with Consortium Agreement. Those that use the data (as opposed to any resulting manuscripts) shall cite and annotate it as follows: *'The data were created by the APPROACH project, funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101120397. For reuse of this data, please, contact APPROACH Consortium. Include your proposed use of the data to assist us in determining your eligibility and to help us navigate possible conflicts between research projects. We will provide you with a short data sharing*

agreement for you and your authorized institutional official to sign prior to your receiving of the data'. This information must be described also in the metadata.

#### 4.4 Protection of personal/special categories of data

With regard to the collection and processing of personal data that will be consolidated during the Widening actions, communication and other dissemination activities, compliance with the Horizon Europe ethical standards, General Data Protection Regulation 2016/679 of the European Parliament and the Council and national data privacy regulations will be ensured. Informed consent for data sharing and long-term preservation will be included in questionnaires dealing with personal data.

## 5. Dissemination

A beneficiary that intends to disseminate its results must give advance notice to the other beneficiaries of (unless agreed otherwise) at least 30 days, together with sufficient information on the results it will disseminate. Each party shall not include in any dissemination activity another party's results or background (data, know-how or information, IPR) without obtaining the owning party's prior written approval.

Dissemination and exploitation will both be a coordinated and collective activity implemented by the entire consortium and an individual set of actions handled by a specific partner on local level. FORTH, as the dissemination Leader, will be responsible for monitoring all planned and performed dissemination and exploitation activities, so as to ensure that the CA and GA requirements are respected and that the involved parties IPR are adequately protected from unauthorised use or any other kind of misuse. Accordingly, the APPROACH produced dissemination and communication materials will be based on the identification of the respective stakeholders they will address at each time and consider their specific needs and interests. Such products will include for example: Printed and electronic flyers and brochures, presentations and posters, banners, videos, fact sheets and press releases, infographics, storyboards, visual maps and e-news, etc.

Concerning scientific publications, it is anticipated that papers for scientific journals and conferences will be prepared and submitted as soon as the project delivers its first results. The APPROACH partners will be regularly invited to submit publications to some conferences and journals, to generate understanding on the project activities and engage the stakeholders. To maximise the impact of the project in the scientific community, the consortium will pursue open access publications. Any scientific publication will be in agreement with the Intellectual Property Rights, IPR, strategy and avoid any possible disclosure of information submitted to patent applications. Due to the different policies of each member of the consortium Gold and Green model will be considered for open access to maximise

public and researcher access. Participants in the project will also gain access to a) the knowledge arising from work carried out under the joint programme of activities and b) the pre-existing know-how of the other participants when it is necessary for the project development. Except where other conditions are set in the Collaboration Agreement, access rights to generated knowledge and pre-existing know-how shall be granted on a royalty-free basis for the development of the project. Access rights for profit-making / commercial use of the knowledge generated within the network are detailed in the Consortium Agreement. Results that may touch directly upon the commercial interests of the partners will require prior agreement of all project partners and the EC before dissemination. The veto of one partner will be sufficient to prevent publication. The Steering Committee will be the forum where the commercial significance of the results and questions concerning patents, copyrights and communications will be discussed to avoid premature disclosure.

## Conclusions-Implications

This is the most recent version of the APPROACH Data Management Plan and has been released as a public report. The APPROACH DMP is intended to be a “living document,” meaning it will be revisited and revised as necessary throughout the project lifespan to reflect changes in project progression. A new DMP will be released for APPROACH anytime is needed, allowing for all major updates to be quickly captured, and for an up-to-date timetable for future updates.