EDAPHOS

D6.8 Data Management Plan (updated)

WP6 – Task 6.2

February 2026 (M30)

Authors: Michel CHALOT (UMLP), Samy STROLA, Mariana TERREROS LOZANO, Adeline PAUL (LGI)

Funded by the European Union

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Document information

Grant Agreement	n°101112768
Project Title	Advanced mapping, risk assessment and nature-based depollution methods are combined to accelerate the recovery of contaminated soils and ensure that ecological restoration enters mainstream business
Project Acronym	EDAPHOS
Project Coordinator	Prof. Michel Chalot, UMLP
Project Duration	1 September 2023 – 31 August 2028 (60 months)
Related Work Package	WP6
Related Task(s)	Task 6.2
Lead Organisation	UMLP
Contributing Partner(s)	All consortium partners
Due Date	28/02/2026
Submission Date	27/02/2026
Dissemination level	PU

History

Date	Version	Submitted by	Reviewed by	Comments
23/01/2026	V1	M. CHALOT (UMLP)	all partners	all partners contributed to Annex V.
27/02/2026	v2	M. CHALOT (UMLP)	A. PAUL (LGI)	Final submitted

Table of contents

Funded by the European Union

List of tables

Summary

The EDAPHOS Data Management Plan (DMP) follows the Horizon Europe (HE) DMP template that was designed to be applied to any HE project that produces, collects or processes research data. It reflects the status of the data that is collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

This DMP deliverable describes the data management principles and strategies, tools and EDAPHOS data that will be produced as part of the project activities and that are relevant to be included in the DMP. The consortium will also aim at open access when publishing papers and articles.

The DMP is a living document to be updated as the implementation of the project progresses and when significant changes occur.

Keywords

Data management, findable, accessible, interoperable and reusable.

Abbreviations and acronyms

Acronym	Description
CA	Consortium Agreement
CC	Common Creative
CSS	Customer Service and Support
D	Deliverable
DMP	Data Management Plan
DPO	Data Protection Officer
DoA	Description of Action
DOI	Digital Object Identifier
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
HE	Horizon Europe
IPR	Intellectual Property Rights
ISSN	International Standard Serial Number
M	Month (of the project)

Funded by the European Union

NDA	Non- Disclosure Agreement
PII	Publisher Item Identifier
PMO	Project Management Office
PU	Public
ORDP	Open Research Data Pilot
SENS	Sensitive
URL	Uniform Resource Locator
V	Version
WP	Work Package
WPL	Work Package Leader
Gold Open Access	Open access publishing (gold open access) means that an article is immediately provided in open access mode on the publisher or journal's website. Some publishers charge Article Processing Chargers (APCs) to make articles open
Green Open Access	Self-archiving (green open access) means that a published article or the final peer-reviewed manuscript is archived (deposited) in an online repository before, alongside or after its publication. In some cases, the author can choose to delay access to the article (embargo period). HE rules state that embargo periods cannot exceed six months, except for publications in social science and humanities where the maximum embargo period is twelve months
Personal Data	Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data. Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the law. Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible
Zenodo	Zenodo is a catch-all research data repository that enables researchers, scientists, EU projects and institutions to share research results, make research results citable, and search and reuse open research results from other projects. Zenodo is harvested by the OpenAIRE portal and hosted by the CERN cloud infrastructure

1 Introduction

1.1 Why would I want to read this deliverable?

This deliverable provides an easy overview of research data the EDAPHOS project is expected to generate, the types and formats of this data, and how this data is processed and stored to make them findable, accessible, interoperable and reusable, according to the principles of FAIR data management. The purpose of the DMP is to contribute to good data handling during the project's lifetime, and to describe how such data will be curated and preserved.

1.2 Intended readership / users

Internally in the project:

- All project participants who are responsible for, or in any way involved with, data collection and data handling can use this document, for instructions on how to handle, store and process data.
- All project participants can use this document to get an overview of all data collected in the project and how this is processed and stored.

External audience:

- All relevant stakeholders who are interested in EDAPHOS related activities and research topics can use this document to get an overview of the data collected in the project, how to access this data, and, if applicable, how to re-use this data in their own activities.
- All persons who voluntarily participate in the pilots and contribute data to the project can use this document to learn how the project processes and stores their data.

1.3 Objectives and scope of the document

The DMP describes the data management life cycle for the data to be collected, processed and/or generated by EDAPHOS project, as a HE project. The DMP aims at defining the management strategy of data generated during the project with the purpose to making research data findable, accessible, interoperable and re-usable (FAIR). The DMP addresses the following points:

- The handling of research data during and after the end of the project,
- What data will be collected, processed and/or generated,
- Which methodology and standards will be applied,
- Whether data will be shared/made open access, and
- How data will be curated and preserved (including after the end of the project).

1.4 Structure of the deliverable

The document is structured following the guideline of HE programme on FAIR Data Management including the following information:

- DMP guiding principles and strategy
- Description of EDAPHOS type of data
- Description of FAIR DATA characteristics including DMP Review Process & data inventory
- Data Security
- Ethical considerations
- Allocation of resources
- Conclusions

Funded by the European Union

1.5 Version

According to the EU's guidelines regarding the DMP (European Commission, 2016), the document may be updated - if appropriate - during the project lifetime (in the form of deliverables). The minimum requirement is that the DMP be updated for each periodic evaluation of the project. If there are none, such an update must be made in time for the final review at the latest.

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.

DMPs should, therefore, have a clear version number and include a timetable for updates.

2 Data Summary

2.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?

Partner	Purpose data that will be collected/generated for the project
UMLP	• UMLP will generate primary data for soil and vegetation in WP2 and WP3. These data will be shared with some of the project partners involved in the other WPs. • UMLP will utilize the data for the deliverables under its responsibility, as well as related project results, open-access publications and/or dissemination activities. • Publication on results, methodologies applied and their implementation on typical use case are foreseen.
CRES	• Generation of primary data from WP3 (soil and crops) • Support the elaboration of the corresponding deliverables in WP3 • Support the activities in WP1 and WP2 • Publication on results, methodologies applied and their implementation on typical use cases are foreseen.
UNIBO	• Generation of basic planta and soil data in WP3 • Support the elaboration of the corresponding deliverables • Support the activities in WP1 and WP2 • Publication of scientific results and methodologies
CSIC	• Generation of primary data from WP3, and some of WP2. • Support the monitoring and management for WP1 and WP2. • Publication on results, methodologies applied and their implementation on typical use case are envisaged.
LGI	• No plan to collect project results and/or data, apart from supporting the monitoring and management of the project itself. • The C&D data collection will be used to collect information on the C&D activities carried out in the project. This data collection tool will facilitate the planning and timing of news and publications that will be visible on the project's communication channels. Basically, it will allow us to analyse the data through KPIs measurements aiming to evaluate the impact of the project through the production of a report.

	• Clustering activities will collect data from related projects and assistants to relevant project events. This data will be used to facilitate contact with the relevant projects and clustering event assistants, and to keep a register of the public reached.
PHYTOWELT	• Generation of primary data from WP3 concerning the distribution of cuttings • Monitoring of data for WP1, WP2 and WP3 concerning the success of phytoremediation and growth of poplar cuttings. • Publication on results, methodologies applied and their implementation on use of poplars for phytoremediation are envisaged.
AMPHOS 21	• Collection of primary data from living labs generated by the partners to incorporate them to a centralized database for visualization and analysis purposes. • Generation of secondary data (geotiff, raw data) resultant from AI/ML models • Publication on results, methodologies applied and their implementation on typical use case are envisaged.
ONERA	• Vectors or geodesic shapes. • Spectral signatures included in spectral database. • Hyperspectral images for the CS2, CS3 and CS6 provided by calibrated a hyperspectral camera installed on an aircraft device (standard raster format compatible with tools widely used, i.e. QGIS, ENVI). • Spaceborne hyper/multi-spectral images for each CS. • Biophysico-(chemical) parameter maps derived from hyper/multi-spectral images.
EVOTROPIA	• Use of only secondary data –either from field experiments or AI/ML generated- as inputs from WPs 1-3 to the quantitative models structured in WPs 4-5. • Although EVOTROPIA will not be generating primary data, it will be monitoring the works of respective WPs and contribute to their formation and classification to ensure compatibility and usability for WPs 4-5. • Use of the data for the deliverables under its responsibility, as well related open-access publications and dissemination activities. Access level will be determined ad-hoc by the agreed confidentiality level.
MICROFLUIDICS	• Generation of Data in WP2 regrading ecotoxicity assessment. Data generated at MIC will be mainly resulting from technical specifications and testing of the microfluidic system to determine the best architecture of the system.
INERIS	• Primary data for the soil ecotoxicity and soil function activities characterization in WP2. These data will be shared with some of the project partners involved in the other WPs. • Use of secondary data collected and/or generated in WPs 1-3 and from sub-contracting laboratory for the ecological risk assessment of CS in WP2. • Use of the data for the deliverables under its responsibility, as well as related project results, open-access publications and/or dissemination activities. • Development of specific templates for the environmental data collection and the ecological risk assessment.

Funded by the European Union

	• Publication on results, methodologies applied and their implementation on typical use case are envisaged.
GIG	• Spatial data regarding previous, current and future land use to evaluate ecosystem service changes (e.g. Corine LAND Cover Data, Urban Atlas, Urban planning) in WP1 and WP2 • Data necessary for calculating the value of ecosystem service indicators (e.g. satellite images, soil properties, climate data,) in WP2 • Ecotoxicological tests to gather the most relevant data regarding test site in WP3 • Market data for evaluation economic value of the ES (e.g. price of electricity from photovoltaic farm, CO2 emissions cost) in WP 4

Table 1: Purpose of data collection and generation for the project by partners.

For more information and guidelines see Annex I, 8.1 Data Collection and 8.2 Data Description.

2.2 What types and formats of data will the project generate/collect?

Partner	Type	Format
UMLP	• Reports and standard documents (deliverables, minutes, templates, presentations, etc.). • Both published and generated data on analytical and biological measurements from experimental activities in WP2 and WP3. • Data will be processed using instrument-specific software and Excel software. • Statistical software will be used to aid interpretation of data. • Dataset, metadata and study results will be summarized in Excel files format, text format, pdf format and PowerPoint format.	.docx .xlss .pptx .pdf .csv Various graphical formats
CRES	• Own generated data from experimental activities in WP3 (pot and field data) • Reports and standard documents (deliverables, templates, presentations, etc.)	.docx .xlss .pptx .pdf Various graphical formats
UNIBO	• Own generated data from experimental activities in WP3 (pot and field data) • Reports and standard documents (deliverables, minutes, templates, presentations, etc.)	.docx .xlss .pptx .pdf Various graphical formats
CSIC	• Data from own analysis of soil and plants from CS 3 related with all activities of WP3. Provide information and samples for WP1 and WP2.	.docx .xlss .pptx .pdf Various graphical formats

	• Reports and standard documents (deliverable, templates, presentations, etc.).	
LGI	• Reports and standard documents (deliverables, minutes, templates, presentations, etc.). Chart tables for the project monitoring (project progresses, financial tables, etc.). • Consortium contacts and mailing lists. • Visuals for C&D activities. • Registration form for meetings/workshops through https://evenium.events/ or similar tool (temporary use)	.docx .xlss .pptx .pdf Various graphical formats
PHYTOWELT	• Reports and standard documents (deliverables, minutes, templates, presentations, etc.). • Chart tables for the project monitoring (project progresses, financial tables, etc.).	.docx .xlss .pptx .pdf Various graphical formats
AMPHOS 21	• Vector • Spectral signatures • Hyperspectral image • Reports and standard documents • Spaceborne image • Produced maps • Raw ASCII data from sensors	ESRI shapefile ASCII or .xlss Raster format with ASCII header docx. xlss .pptx .pdf Native format GeoTiff or raster format with ASCII header Images .png .jpg
ONERA	• Vector • Spectral signatures • Hyperspectral image • Reports and standard documents • Spaceborne image • Produced maps	ESRI shapefile ASCII or .xlss Raster format with ASCII header docx. xlss .pptx .pdf Native format GeoTiff or raster format with ASCII header
EVOTROPIA	• Deliverables of any kind (reports, prototypes, methods, digital tools, videos). • Spreadsheets with time-series (e.g. temperature, ET and generally weather/climate), cross-sectional data (e.g. 7 CSs) or panel data as a coordinated combination of both (i.e. same variables, time scales and periods) for econometric analyses and Soil Observatories' databases formation.	.docx .pdf .pptx .xls .csv Any picture or infographic containing flow designs or statistical data of a template able to process (e.g. .ai) Any video format with preference for high device

Funded by the European Union

	• Derived data with special 1st-level classification, such as 5-stage life-cycle data that are further structured at 2nd-level in spreadsheet form (see above).	compatibility (e.g. .mp4, .mov) and secondarily for compression ability (e.g. .wmv)
MICROFLUIDICS	• Reports and standard documents (deliverables, protocols, presentations). • Data from technical developments (images, protocols, videos).	.docx .pptx .csv .xls .pdf
INERIS	• Reports and standard documents (deliverables, minutes, templates, presentations, etc.). • Both published and generated data on analytical and biological measurements. • Data will be processed using instrument-specific software and Excel software. Statistical software will be used to aid interpretation of data. • Dataset, metadata and study results will be summarized in Excel files format, text format, pdf format and PowerPoint format.	.docx .xlss .pptx .pdf .csv Various graphical formats
GIG	• Both published and self-generated data on analytical and biological measurements (pictures/imagines/graphs/tables/etc.). Dataset and study results will be summarized in excel files format, as well as text or pdf format. • Both public and generated data provided by the consortium partners. • Reports and standard documents (deliverables, minutes, templates, presentations, etc.).	.docx .xls .pptx .pdf Various graphical formats (.jpg; .png; .GeoTIFF, etc.)

Table 2: Types and formats of data generated/collected by partners.

For more information and guidelines see Annex I, 8.1 Data Collection and 8.2 Data Description.

2.3 Will you re-use any existing data and what will you re-use?

Partner	Existing data that will be used or re-used
UMLP	Any existing data from previous projects (EU or national project) that fit with the management of contaminated soil purpose. Especially historical data from case studies concerning soil analytics (chemistry, biology...) and vegetation (metal accumulation).
CRES	No data from previous works will be utilized in this project. Re-use of project monitoring dashboard's templates.
UNIBO	No data from previous works will be utilized in this project. Re-use of project monitoring dashboard templates.

CSIC	No data from previous works will be utilized in this project. Re-use of project monitoring dashboard's templates.
LGI	No data from previous works will be utilised in this project. Re-use of project monitoring dashboard's templates.
PHYTOWELT	Existing data from previous projects under the form of analytical background information.
AMPHOS 21	No data from previous works will be utilized in this project.
ONERA	No data from previous works has been identified to date to be re-used in this project. ONERA will use its own software listed in the CA to process data.
EVOTROPIA	Due to lack of structural comparability (i.e. different, hence incomparable variables) or even complete inexistence (WP 4-5 models have not been applied so far in NBS), re-use of previous/existing data is not expected. Across the project's deployment this position will be periodically re-examined.
MICROFLUIDICS	Data obtained in the scope of previous similar developments will be reused, as long as they fall within the scope of EDAPHOS.
INERIS	Any existing data from previous projects (EU or national project) that fit with the environmental risk assessment purpose. Especially data concerning soil analytics (chemistry, biology...).
GIG	An existing data in the form of spatial data will be used (topographic data, satellite images, land cover maps, etc.). These data will constitute the base material, in further work and analysis under WP1. Under WP4 due to the specificity of the analysis requiring the development, adaptation and/or modification of existing methodologies individually for each CS, re-use of previous/existing data is not expected.

Table 3: List of data that will be used or re-used by partners.

For more information see Annex I, 8.1 Data Collection.

2.4 What is the origin/provenance of the data, either generated or re-used?

The EDAPHOS project will generate new data from remote sensing acquisitions (WP1 partners *i.e.* ONERA, GIG), field and laboratory experiments (WP2 & WP3 partners; *i.e.* UMLP, GIG, CRES, UNIBO, INERIS, PHYTOWELT, MICROFLUIDICS) and desktop research and communication activities (WP4-5-6 activities, *i.e.* LGi, EVO, GIG, AMPHOS21, UMLP). Some data will come from already existing and available databases. The conditions of reuse of obtained data from these databases will depend on the discussion between the EDAPHOS project and each data provider. In the table below, each partner engaged in the obtention of data details the origin and provenance of data:

Partner	Origin of data that will be generated or re-used
UMLP	Re-used data from previous R&D projects on samples obtained through field campaigns in previous projects. New data will be generated from new samples obtained along the project period from the two case-studies Vieux Charmont, and Carrières sous Poissy.

Funded by the European Union

	New data will be generated based on desktop research and communication activities with other initiatives that are developing complementary activities (through e.g. other Horizon projects). The data used will be shared by and with the consortium partners.
CRES	The data will be generated by the experimental activities carried out in the case studies; Kozani (north-west Greece) and Lavrion (near Athens)
UNIBO	The data will be generated by the experimental activities carried out in the Castel vetro site (CS5)
CSIC	The origin of data are from the CS 3, in the Río Tinto mine area, from its soils, own plants and poplar hybrids and co-crops that we will stablish in the area.
LGI	The data used will be shared by and with the consortium partners.
PHYTOWELT	Re-used data from previous R&D projects on samples obtained by field campaigns in previous projects.
AMPHOS 21	Collected from living labs from other partners in the context of the project.
ONERA	New data acquired and new maps generated in EDAPHOS by existing software developed by ONERA in previous studies. Part of existing data and information will originate from: spatial information systems, public available portals and programs (e.g. Corine Land Cover, Google Maps, OpenStreetMap, etc.). It is possible that other data will be used as the research project advances, but these sources and types and not currently foreseeable.
EVOTROPIA	Whether the data received from WPs 1-3 as inputs to WPs 4-5 are existing or generated for the first time in EDAPHOS via real observations or AI/ML simulations, will be shared with consortium partners and will be strictly used for the project's purposes unless agreed otherwise.
MICROFLUIDICS	Data will be either re-used with permission (data obtained in the same institution, with complete freedom to operate) or they will be generated during technical developments in the scope of EDAPHOS.
INERIS	New data will be generated from new sample obtained along the project period. Biological data (mainly ecotoxicity data and ecological data) will be generated according to (or based on) OECD or ISO protocols. Some ecological data will be generated by sub-contracting laboratory (i.e. nematofauna biodiversity, soil DNA measurement for microbial diversity analyses...). Re-use data will come from in-house research during previous project or programs, field campaigns in previous projects involving other partners, data generated by other partners in the WPs 1-3 of the current project.
GIG	The generated data (new and re-used) come from in-house research facilities and laboratories or were generated by other project partners. If analyses are outsourced, the origin of the data is identified. Part of the data will be generated using measurement technique/ equipment (e.g. unmanned aerial vehicles (drones), spectral measurements, etc.). Raw data will be processed and interpreted, using software dedicated to the devices. Part of existing data and information will originate from: spatial information systems, public available portals and programs (e.g. Corine Land Cover, Google Maps, OpenStreetMap, etc.). It is possible that other data will be used as the research project advances, but these sources and types and not currently foreseeable.

Table 4: Origin of data generated or re-used by partners.

For more information and guidelines see Annex I, 8.4 Data sharing.

2.5 What is the expected size of the data that you intend to generate or re-use?

Partner	Size of data generated or re-used
UMLP	Most of the generated data will be raw and processed data (bigger data sizes), however most of it will not be disseminated in general, but only compared with other laboratories. The transfer of the data, however is still easy to handle. Secondary data (e.g. reports) is mainly text based and showing statistics, easy to distribute and the estimated sizes are rather small. No additional costs need to be estimated here for data transfer and storage. Relatively small file sizes that are KB and MB, but some could extend to GB.
CRES	Data will be mainly numerical and text based, presentation and visuals. The overall size is expected to be in the range of MB to GB.
UNIBO	Data will be mainly numerical and text based, presentation and visuals. The overall size is expected to be in the range of MB to GB.
CSIC	The data generated will be in various excels, graphic and text formats. The total size is expected to be in the GB range.
LGI	Data will be mainly text based, presentation and visuals. The overall size is expected to be in the range of GB.
PHYTOWELT	<10 GB
AMPHOS 21	In the range of GB. An initial estimation ranges from 20 to 200GB.
ONERA	The size of each hyperspectral image acquired by aircraft device in EDAPHOS represents 5 to 10 Go. The total amount could be higher than 100 Go.
EVOTROPIA	Ideally, spreadsheet data as inputs to WPs 4-5 will concern at least three years for at least monthly basis (preferably weekly) to achieve a “critical mass” for higher econometric modelling reliability (at least 36-156 observation points). Although these data in .csv or .xls formats are of low size with the inclusion of formulas and macros, as well as all the addition of all other mentioned formats the overall data size is expected to be in the GB range.
MICROFLUIDICS	Data will be mainly text based, presentation and pictures. The overall size is expected to be in the rage of GB.
INERIS	The data acquired and generated will be in various graphic, photos, text, presentation formats. The overall size is expected to be in the range of GB (1 to 3 Go).
GIG	The data acquired and generated will be in various graphic and text formats. The total size is expected to be in the GB range.

Table 5: Size of data generated or -reused by partners.

For more information and guidelines see Annex I, 8.3 Data volume.

Funded by the European Union

2.6 To whom might your data be useful ('data utility'), outside your project?

Please see groups of possible stakeholders below:

Sector	Stakeholders
Projects/Initiatives	EU funded Projects Other grant funded projects Initiatives
Industry	Plant and seed producer Biofertilizer producer Site manager Consulting engineering in environment
Investors	Accelerators Angels Corporates Venture Capital Private Equity Government non-profit Advisor Banks Peer-to-peer investors
EU/MS governments, policy makers and institutions	Ministries Municipalities EU institutions & Regulators Environmental agencies National and regional governments and regulators/authorities International organizations
Research peers: academic sector	Universities Research organizations Standardization and certification bodies
General public	Consumer associations Citizens Other
Other: consultancies, NGOs, media	Consultants NGOs Media Law enforcement agencies National or federal police Forensic investigators

Table 6: List of main sub-groups of stakeholders.

For more information and guidelines see Annex I, 8.4 Data sharing.

3 Data Management Policy: FAIR Data

3.1 Making data findable, including provisions for metadata

Digital Object Identifier (DOI)

All open data, publications and open source software produced in EDAPHOS will be identifiable and locatable by means of a persistent Uniform Resource Locator (URL). If possible, open EDAPHOS results will be assigned a Digital Object Identifier (DOI) in order to make content easily and uniquely citable.

Open results that are deposited in the default Open Access repository (Zenodo, see below) will be assigned a DOI automatically and will benefit also from Zenodo's DOI versioning support.

Open results that are deposited in institutional repositories, repositories of scientific publishers or other data and research repositories will be at least definable by a persistent URI. If the institution is a DOI registrant that has an agreement with a DOI registration agency, a DOI will be assigned, too.

Whether scientific publications will be assigned a unique identifier like DOI, Publisher Item Identifier (PII), International Standard Serial Number (ISSN), etc. depends on the open access strategy (green or gold) chosen by the editors and thus also on the respective scientific publisher and the chosen research repository.

Naming conventions

Files and folders at data repositories will be versioned and structured by using a name convention consisting as follow (see D7.1):

EDAPHOS_[name of the document]_Vxy_date_[partner acronym/person name].FileExtension

FileType are:

- D stands for Deliverable
- DS stands for DataSet
- F stands for File (generic, ex. images, table, document)
- Other acronyms will be defined upon specificities

For specific files (e.g. CS – images, maps, …), the naming convention can deviate from the above one with the aim to identify them easier.

Zenodo

EDAPHOS open data will be collected in an open online research data repository: Zenodo. Its repository structure, facilities and management are in compliance with FAIR data principles. Zenodo is an OpenAIRE that allows researchers to deposit both publications and data, providing tools to linking them to these through persistent identifiers and data citations. Zenodo is set up to facilitate the finding, accessing, re-using and interoperating of datasets, which are the basic principles that ORD projects must comply with. Zenodo repository is provided by OpenAIRE and hosted by CERN. Zenodo is a catch-all repository that enables researchers, scientists, EU projects and institutions to:

- Share research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science;
- Display their research results and get credited by making the research results citable and integrating them into existing reporting lines to funding agencies like the European Commission;
- Easily access and reuse shared research results;
- Assign a DOI automatically to each result deposited in the repository,
- Integrate their research outputs with the OpenAIRE portal.

Search keywords

Zenodo allows to perform simple and advanced search queries on Zenodo using the keywords. Zenodo also provides a user guide with easy-to-understand examples. The Data Controllers at each pilot site will be responsible for uploading public datasets that they have generated and to assign specific keywords relevant to these datasets. Dataset specific keywords must be

Funded by the European Union

descriptive to the content of the dataset. In addition, the project has defined a set of general keywords that should apply to all public datasets, scientific publications and public deliverables. These are as follow:

- Critical Raw Materials
- Digital Product Passport
- Traceability
- Chain of custody
- Verifiable credentials
- Blockchain

Version numbers

Individual file names and datasets will contain version numbers that will be incremented at each revision (V*xyz*). For publications, versioning is in general not necessary.

Zenodo provides DOI versioning of all datasets uploaded to their communities, which allows us to edit and update the uploaded datasets after they have been published. This also allows us to cite specific versions of an upload and cite all versions of an upload.

Metadata

Metadata is data on the research data themselves. It enables other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers, can better determine whether the dataset is relevant and useful for their own research. Metadata will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data.

All metadata will be made available and is licensed under CC0, as soon as the different topic specific metadata schemas are established within the project. It is currently recognized within the project, that there are various gaps and conflicting metadata schemas and definition, which will require first a harmonization or possibly a distinction between the different fields.

All datasets, which will be shared openly on trusted repositories will be available for at least 10 years. This will be one of the criteria, which is in the checklist to assess any repository, before it is put on the list of trusted repositories, which should be used by project partners. Furthermore, it will be checked if the metadata will remain openly available, even after the datasets have been discarded after the data retention period has expired. This also depends on the arrangement of the repository to warrant functionality over a given period of time. Assessing this part will also be part of the checklist to assess repositories. This common repository checklist will be part of the next version of the DMP.

The following deposition metadata fields are mandatory for EDAPHOS:

- the terms “European Union (EU)” and “Horizon Europe”;
- the name of the action, acronym and grant number;
- the title and description of the deposition metadata,
- the upload type (publication, dataset, software, …),
- the publication date (ISO8601 format, YYYY-MM-DD), and length of embargo period if applicable;
- the creators/authors of the deposition,
- the persistent identifier (DOI),
- the used keywords.

This minimal metadata schema can be extended by arbitrary subjects upon request of the consortium.

For more information and guidelines see Annex I, 8.4 Data Sharing and 8.5 Metadata & documentation.

3.2 Making data openly accessible

3.2.1 Data made openly available as the default

In order to maximise the impact of EDAPHOS research data, the results are shared within and beyond the consortium. Selected data and results will be shared with the scientific community and other stakeholders through publications in scientific journals and presentations at conferences, as well as through open access data repositories.

The EDAPHOS project datasets are first stored and organized in a database by the data owners (personal computer, or on the institutional secure server) and on the project database (project website). All data are made available for verification and re-use, unless the task leader can justify why data cannot be made openly accessible. To protect the copyright of the project knowledge, Creative Commons license will be used in some cases.

The EDAPHOS dataset deliverables are both public (data access policy unrestricted) and they will be accessible by:

- EDAPHOS project web site
- Partners database
- OpenAIRE
- Zenodo for ORDP data and datasets
- Open access journals

All data deposited on Zenodo are accessible without restriction for public.

3.2.2 How will the data be made available

EDAPHOS open results will be made available according to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon Europe.

Open data: All open results (data, software, scientific publications) of the project will be openly accessible at an appropriate Open Access repository (i.e. Zenodo) as soon as possible. Specifically, research data needed to validate the results in the scientific publications will be deposited in a data repository at the same time as a publication. Non-public research data will be archived at the repository using a restricted access option.

Scientific publications: Providing open access to peer-reviewed scientific publications can be ensured either by publishing in green or gold open access journals with or without author processing fees. Any scientific publications from PROJECTNAME and the related bibliographic metadata must be made available as open access and announced on the project website (https://www.EDAPHOS.eu/) as well as in the OpenAIRE portal (https://www.OpenAIRE.eu/) and the R&I Participant Portal (https://ec.europa.eu/research/participants). To automate the process of reporting scientific publications and related research data in OpenAIRE, the publication should be deposited in an OpenAIRE-compliant repository, either by the authors of the publication (green open access) or by a scientific publisher (gold open access). While additional forms of disseminating open access papers, including academic social network sites such as ResearchGate (https://www.researchgate.net/) are possible, an electronic copy of the publication has to be deposited in suitable open access repository in the first place. According to the European Research Council's Guidelines on Open Access, "Venues such as Research Gate or Academia.edu that require users to register in order to access content do not count as repositories. The posting of publications on a personal, institutional or project specific webpage or the deposit in a publicly accessible Dropbox account is not sufficient to satisfy the requirements either."

If the chosen repository is not fully OpenAIRE complaint, the publications or data must be linked at https://www.openaire.eu/participate/claim with the respective funding agency (European Commission). Green open access journals or gold open access journals without author

Funded by the European Union

processing fees should be preferred for disseminating scientific publications of the PROJECTNAME project. Nevertheless, the journal's visibility and prestige (translated in the Impact Factor) of the journal, together with the speed of publication, should be considered when choosing a journal for publication of a manuscript. According to the EC recommendation, authors of the publication are encouraged to retain their copyright and grant adequate licences to publishers.

- Green open access (self-archiving): Green open access or self-archiving means that the published article or the final peer-reviewed manuscript is archived by the researcher itself in an online repository, in most cases after its publication in the journal. The journal must grant the researcher the permission to self-archive the final peer-reviewed article, at the latest, 12 months after publication. For finding suitable green open access publishers, researchers are encouraged to consult RoMEO (http://sherpa.ac.uk/romeo), a searchable database of publisher's policies regarding the self- archiving of journal articles on the web and in Open Access repositories.
- Gold open access (open access publishing): Gold open access means that the publication is available by the scientific publisher as open access. Some journals require an author-processing fee for publishing open access. Author-publishing fees for gold open access journals can be reimbursed within the project period and budget. Some publishers allow the researcher to deposit a copy of the article in a repository, sometimes with an embargo period. For finding suitable gold open access publishers, researchers are encouraged to consult the Directory of Open Access Journals (https://doaj.org/), a service that indexes high quality, peer-reviewed open access academic journals that use an appropriate quality control system.

3.2.3 Methods and/or software needed to access the data

Regarding the mere access to open data deposited as data files in a data repository, there are no special methods or software tools needed. The files can be downloaded from the data repository using a standard web browser. The offline viewing, interpreting, processing and editing of data files downloaded from the data repository, it heavily depends on the type and format of the data.

3.2.4 Data sharing exceptions

1. Copyright and permissions for re-using third-party data sets: Processing and combining input data from many different sources may lead to unclear IPR situations regarding the generated output data, therefore such repurposed data (e.g. model output data) can only be made open if any of the underlying data (e.g. model input data) is open, too. For this: potential users must contact the IPR (Intellectual Property Rights) team or the data owner in order to gain access. If necessary, appropriate IPR procedure (such as non- disclosure agreement - NDA) will be used.
2. Personal data treatment and confidentiality issues: Datasets referring to the quality and quantity of certain elements at risk, such as people and critical infrastructures, are not open by default as their publication may pose privacy, ethical or security risks.
3. Data-driven business model: Data that will be exploited commercially will not be made open.

3.2.5 Specific partner provisions

There are no specific provisions for any of the partners.

3.2.6 Sensitive data restrictions

All partners: Each partner, who will institute a restricted access to their datasets, will setup a procedure that will identify persons who require access to these datasets. This identification and access procedure will be described by the respective partner in detail in the next version of the DMP. The RDM of the project will review the procedure and can make recommendation. In the

next version of the DMP all the partners with restricted access will be listed and their procedures will be described.

3.2.7 Data access committee

Each partner, which has personal and/or sensitive data, has to setup a procedure, which identifies persons, who require access to these datasets. Besides that, they have to have a data access committee, which involves at least one legal specialist in personal and sensitive data, which grants access to the data based on the previously defined procedure. In the next version of the DMP all the partners with personal and/or sensitive data will be listed, and their procedures will be described. This also includes the data access committee.

3.2.8 How will access be provided if there are restrictions on use to data

Where a restriction on open access to research data is necessary, attempts will be made to make data available under controlled conditions to other individual researchers. In the case where restricted or embargoed data is stored in the Zenodo repository, information about the restricted data will be published in the repository, and details of when the data will become available will be included in the metadata. According to the Q&A session "Open Research Data in HE and Zenodo repository", Metadata for both open, closed, embargoed and restricted records are always publicly available in Zenodo. Data files and data sets for restricted access records are only visible to their owners and to those the owner grants access. Restricted access allows a researcher to upload a dataset and provide the conditions under which he/she grants access to the data. Researchers wishing to request access must provide a justification for how they fulfil these conditions. The owner of the dataset gets notified for each new request and can decide to either accept or reject the request. If the request is accepted, the requester receives a secret link which usually expires within 1-12 months.

For more information and guidelines see Annex I, 8.4 Data sharing

3.3 Making data interoperable

General data and metadata vocabularies, standards, formats and methodologies will be followed. Nevertheless, there are no generally accepted and standardized field specific data and metadata vocabularies, standards, formats and methodologies. Partners will try to follow community-endorsed interoperability best practice. But first it has to be verified that the majority of the project partners are technically capable implementing these best practices. More details will be given in the next version of the DMP.

EDAPHOS will have to generate project specific ontologies and vocabularies. Partners will try to follow the examples of already existing ontologies and vocabularies to achieve some harmonization. But in any case, will be openly published as soon as a first version has been approved. They will be updated regularly. The next version of the DMP will deliver updated information concerning this topic.

3.4 Increase data re-use

The research data management team will advise partners in the project concerning research outputs other than data. The team will develop a process how to handle the research output generated within the program. This process will be described in more detail in the next version of the DMP.

For more information and guidelines see Annex I, 8.6 Storage and security

Funded by the European Union

3.5 DMP quality and review process & data inventory

Internal process of quality evaluation and reporting is activated throughout the entire project duration to assess both project data /products and project process (See the D6.4 - Project Quality Plan). An internal peer review is performed for the main project deliverables to guarantee the deliverable is developed with a high level of quality. Each WPL has to submit all the produced documents to another partner assigned as internal reviewer to check for the quality of the documents produced.

Results data will be also analysed and collected throughout the project entire duration. To this purpose the Dissemination and Communication Plans (D5.1) will also be filled in by each partner about every six months: it includes the description of articles, papers and scientific publications too. Thus, all research data generated and publications will be analysed and described by using the Data Inventory Table (Annex II), WPLs and partners authors of publications are required fill in periodically.

Further updating of the DMP will include the eventually updating of online research data repository where data are collected and shared and the data the description of dataset and research data gradually generated and collected.

3.6 Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g., software, workflows, protocols, models, etc.) or physical (e.g., new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

4 Data security

4.1 Data access

The following guidelines will be followed in order to ensure the security of the data:

- Store data in at least two separate locations to avoid loss of data;
- Encrypt data if it is deemed necessary by the participating researchers;
- Limit the use of USB flash drives;
- Label files in a systematically structured way in order to ensure the coherence of the final dataset.

All project deliverables and data will be stored and shared in the dedicated Teams repository restricted to the project consortium. As an initial step, only the Consortium Partners will have access to the repository where dataset and metadata are filed. The protection of data will be ensured through procedures and appropriate technologies (e.g., HTTPS protocol for the encryption of all internet transactions and appropriate European and Internet security standards from ISO, ITU, W3C, IETF and ETSI). If data will be kept in a certified repository, then the security standards of that repository will apply.

Following, scientific publications and articles, the dataset deliverables and the final demonstrator research results will be shared through Zenodo and other database to promote the data making FAIR.

4.2 Data preservation

Data preservation refers to data upkeep and maintenance to ensure that the integrity of the data is upheld in the future. This includes properly maintaining the data repository and data backups to ensure the long-term value of the data.

Data backups are expected to occur once a week, though this is subject to change based on the amount of data that will ultimately need to be backed up. Estimating the size of the data is difficult to accomplish at this time since the data that will be initially included in the project has not been fully identified yet. An initial inventory will be identified by the Consortium. A better impression on the size of the data can be made after these datasets have been identified. However, EDAPHOS is designed and intended to have useful data added on a consistent basis, so the size will likely change over time, as well.

The Teams repository has no limits on the size of data that can be stored.

The EDAPHOS data can be downloaded and copied to personal computers without monitoring by each partner having access to the project repository.

5 Ethical aspects

5.1 General Data Protection Regulation (GDPR)

The EDAPHOS project is fully compliant with the General Data Protection Regulation (GDPR) laid out in Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC and respects regulations on intellectual property rights (IPR).

Repository and database access will be fully anonymous. If a user wishes to use the Teams of the EDAPHOS project, they will be required to register an account. The user must provide their email address and grant their consent that their email address be processed and used for account-related communications and management. The user is free to revoke their consent at any time, which would lead to the deletion of their access to the Teams.

5.2 Sensitive data

Sensitive data is data that is either private or confidential and includes personal user data. The proper management of sensitive data is imperative to maintain the individual privacy and remain in compliance with both EU and international regulations.

In order to ensure sensitive data is properly managed, data that is considered sensitive should first be identified. Thus, the main ethical and privacy issues with sensitive data arise from ensuring the data remains private and that proper consent is obtained before the data is shared or published in any way. Sensitive data will be stored in privately by partners handling the data. Measures to protect the privacy of individuals providing sensitive data will be taken in any instance where sensitive data will be collected and published. Furthermore, when partners are meant to deal with

Funded by the European Union

sensitive data they must notify the Consortium and comply with the Joint data controller guidelines (see annex III) and complete the table under annex IV.

6 Allocation of resources

Costs

Costs related to open-access to research data in HE are eligible for reimbursement under the conditions defined in the GA, but also other articles relevant for the cost category chosen. Project beneficiaries will be responsible for applying for reimbursement for costs related to making data accessible to others beyond the consortium.

The costs for making data FAIR includes:

- Fees associated with the publication of scientific articles containing project's research data in "Gold" Open access journals. The cost sharing, in case of multiple authors, shall be decided among the authors on a case-by-case basis;
- Project Website operation: to be determined;
- Data archiving at Zenodo and on other on line data base: free of charge
- Copyright licensing with Creative Commons: free of charge.

Each partner is responsible for the data they produce. Any fee incurred for Open Access through scientific publication of the data will be the responsibility of the data owner (authors) partner(s).

No immediate costs are anticipated for open data that is stored for long-term preservation in the Zenodo repository. Additional details will be reported, as needed, in future versions of the DMP.

Responsibility

Data management activities concern the whole project and needs to be coordinated and monitored both at project and work package level. Data management is also linked to publication of project results and thus dissemination activities.

The overall responsibility for data management lies with the project coordinator, Daniel Monfort from UMLP.

The project member of the Executive Committee (i.e. WPLs) are also responsible of the Data Management of EDAPHOS dataset and research data.

The Coordinator (project data manager) and the WPLs (WP data managers) are responsible for:

- Developing and implementing the data management plan and policy in cooperation with the project management,
- Monitoring data management activities (both collection and publication) and deadlines,
- Monitoring that open results (data and software) are deposited in the default repository or a complementary OpenAIRE-compliant repository and that are linked with EDAPHOS,
- Providing support and sending reminders to all partners,
- Writing and uploading the data management plan,
- Offering assistance in choosing the right publication path (green or gold open access),
- Offering customized help and further guidance for publishing scientific publications,
- Ensuring that the open access policy of the journal complies with the HE open data requirements before the researcher submits a manuscript,
- Monitoring that green access (self-archiving) publications are deposited in repositories and sending reminders to partners,
- Monitoring that metadata about publications is made available in the R&I Participant Portal (preferably automatically through OpenAIRE) and on the EDAPHOS website,

- Monitoring that research data related to a publication is made available in repositories and linked to respective publication,
- Monitoring possible embargo periods and sending reminders to partners,
- Monitoring that publications available in OpenAIRE are properly linked with EDAPHOS,
- Describing the data (by means of appropriate metadata) or scientific publication in accordance to the EDAPHOS data management policy (e.g. according to the chosen metadata standard) and with help of the tools (e.g. template, web form, ...) provided by the project.

For more information and guidelines see Annex I, 8.6 Storage and security, 8.8 Roles and Responsibilities, 8.9 Preservation

7 Conclusion

This document describes the man principles and guidelines for the Data Management for the EDAPHOS project. As living document, it will be updated throughout the project lifetime. Further updating of the DMP will include the eventually updating of online research data repository where data are collected and shared, the standard of data structuration, the description of dataset and research data gradually generated and collected.

Funded by the European Union

Bibliography

European Commission. ***Horizon Europe Programme***. *Guidance for the classification of information in research projects*.

European Commission. ***Horizon Europe Programme***. *Guidelines on FAIR Data Management in Horizon Europe*.

European Commission. ***Horizon Europe Programme***. *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon Europe*.

European Union (2020). Intellectual property rights. https://europa.eu/youreurope/business/running-business/intellectual-property/rights/index_en.htm

European Parliament (2021). *Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32021R0695

European Parliament (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*. https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32016R0679

8 Annex I - Digital Curation Centre (DCC) – Information

This chapter summarizes information provided by the Digital Curation Centre (DCC). The DCC was founded in 2004 to enable research data management the use of data that are findable, accessible, interoperable and reusable (FAIR). The information shared in this chapter can be seen as a guideline for the implementation of a successful Data Management Plan.

8.1 Data Collection

- Outline how the data will be collected and processed. This should cover relevant standards or methods, quality assurance and data organisation.
- Indicate how the data will be organised during the project, mentioning, e.g., naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier to find, understand and reuse.
- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture, data entry validation, peer review of data or representation with controlled vocabularies.
- See the DataOne Best Practices for data quality.

8.2 Data Description

- Give a summary of the data you will collect or create, noting the content, coverage and data type, e.g., tabular data, survey data, experimental measurements, models, software, audio-visual data, physical samples, etc.
- Consider how your data could complement and integrate with existing data, or whether there are any existing data or methods that you could reuse.
- Indicate which data are of long-term value and should be shared and/or preserved.
- If purchasing or reusing existing data, explain how issues such as copyright and IPR have been addressed. You should aim to minimise any restrictions on the reuse (and subsequent sharing) of third-party data.

8.3 Data volume

- Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data, and other secondary outputs (e.g., reports).
- Consider the implications of data volumes in terms of storage, access and preservation. Do you need to include additional costs?
- Consider whether the scale of the data will pose challenges when sharing or transferring data between sites; if so, how will you address these challenges?

8.4 Data sharing

- How will you share the data e.g. deposit in a data repository, use a secure data service, handle data requests directly or use another mechanism? The methods used will depend on a number of factors such as the type, size, complexity and sensitivity of the data.
- When will you make the data available? Research funders expect timely release. They typically allow embargoes but not prolonged exclusive use.

Funded by the European Union

- Who will be able to use your data? If you need to restrict access to certain communities or apply data sharing agreements, explain why.
- Consider strategies to minimise restrictions on sharing. These may include anonymising or aggregating data, gaining participant consent for data sharing, gaining copyright permissions, and agreeing a limited embargo period.
- How might your data be reused in other contexts? Where there is potential for reuse, you should use standards and formats that facilitate this, and ensure that appropriate metadata is available online so your data can be discovered. Persistent identifiers should be applied so people can reliably and efficiently find your data. They also help you to track citations and reuse.

8.5 Metadata & documentation

What metadata will be provided to help others identify and discover the data?

- Researchers are strongly encouraged to use community metadata standards where these are in place. The Research Data Alliance offers a Directory of Metadata Standards. Data repositories may also provide guidance about appropriate metadata standards.
- Consider what other documentation is needed to enable reuse. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data and software used to collect and/or process the data.
- Consider how you will capture this information and where it will be recorded, e.g., in a database with links to each item, in a 'readme' text file, in file headers, etc.

8.6 Storage and security

- Describe where the data will be stored and backed up during the course of research activities. This may vary if you are doing fieldwork or working across multiple sites so explain each procedure.
- Identify who will be responsible for backup and how often this will be performed. The use of robust, managed storage with automatic backup, for example, that provided by university IT teams, is preferable. Storing data on laptops, computer hard drives or external storage devices alone is very risky.
- See UK Data Service Guidance on data storage or DataONE Best Practices for storage.
- Also consider data security, particularly if your data is sensitive e.g., detailed personal data, politically sensitive information or trade secrets. Note the main risks and how these will be managed. Also note whether any institutional data security policies are in place.
- Identify any formal standards that you will comply with, e.g., ISO 27001. See the DCC Briefing Paper on Information Security Management - ISO 27000 and UK Data Service guidance on data security.

8.7 Ethics and privacy

- Investigators carrying out research involving human participants should request consent to preserve and share the data. Do not just ask for permission to use the data in your study or make unnecessary promises to delete it at the end.
- Consider how you will protect the identity of participants, e.g., via anonymization or using managed access procedures.

- Ethical issues may affect how you store and transfer data, who can see/use it and how long it is kept. You should demonstrate that you are aware of this and have planned accordingly.
- See UK Data Service guidance on consent for data sharing.
- See ICPSR approach to confidentiality and Health Insurance Portability and Accountability Act (HIPAA) regulations for health research.

8.8 Roles and Responsibilities

- Outline the roles and responsibilities for all activities, e.g., data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Individuals should be named where possible.
- For collaborative projects you should explain the coordination of data management responsibilities across partners.
- See UK Data Service guidance on data management roles and responsibilities or DataONE Best Practices: Define roles and assign responsibilities for data management.

8.9 Preservation

- Outline the plans for data sharing and preservation - how long will the data be retained and where will it be archived? Will additional resources be needed to prepare data for deposit or meet any charges from data repositories?
- See the DCC guide: How to appraise and select research data for curation or DataONE Best Practices: Identifying data with long-term value.

9 Annex II – Datasets collected in the project

Dataset reference and name: EDAPHOS – WP1 Organisation in charge: ONERA		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Spectral signature	ASCII or .xlss	ONERA software
Vector	Shapefile ESRI	
Image	Raster format with ASCII header, Geotiff or native format (spaceborne image)	ONERA software for aircraft image processing and bio-(physico-)(chemical-) map generation
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
A Few dozen to 100 Go or more	Non-sensitive data will be accessible to the public during and after the project. However, sensitive data will be made available upon request as described in the CA. The data acquired by ONERA will be opened with the agreement of	Owing to the large data size, ONERA could store metadata files on suitable EDAPHOS storage device (i.e. WebGIS, FLEXX or TEAMS-MS, defined in the next) and the images and maps will be store on internal device and share for partners who request them.

Funded by the European Union

	the partner in charge of managing the experimentation site.	
Name of the dataset	Need and gaps in due diligence	
Description	Information that will help project partners to request and use the data.	
Media Type	Text	
Language(s)	English	
Use & re-use	Assessment of needs and gaps in due diligence	
Size	< 1GB	
Format/license	.docx, .xlsx, .pdf, .txt (to define)	
Version number	Not relevant	

Dataset reference and name: EDAPHOS – WP2 Organisation in charge: INERIS		
Data Type	Data Standards - Formats	Data Generation Software
Data outputs such as spreadsheet and reports, methods, protocol	.docx, .xlsx, .pptx, .pdf	MS Office (.docx, .pdf, .pptx,), Adobe Acrobat (.pdf reader or/and editor)
Estimated Data Size	Data Sharing	Storage and Preservation
1 GB	Non-sensitive data will be accessible to the public during and after the project. However, sensitive data will be made available upon request as described in the CA.	WP2 folder FLEXX and TEAMS-MS; WebGIS if useful; internal devices.
Name of the dataset	Need and gaps in due diligence	
Description	The WP2 dataset will contain chemical, ecotoxicological and ecological data generated on each CS as well as ecosystem services and risk assessment results. These data are expected to be useful to partners involved in the other WPs (3, 4, 5).	
Media Type	Text, spreadsheets, presentation, images, graphics	
Language(s)	English	
Use & re-use	Assessment of needs and gaps in due diligence	
Size	1 GB	
Format/license	.docx, pdf, .pptx, .xls., .jpg; .png; etc.	
Version number	Not relevant	

Dataset reference and name: EDAPHOS – WP3 Organisation in charge: UMLP		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Reports, prototypes, methods, spreadsheets with cross-sectional and/or panel data and algorithms, etc.	.docx, pdf, .pptx, .xls., .jpg; .png; etc.	MS Office (.docx, .pdf, .pptx,), Adobe Acrobat (.pdf reader or/and editor)
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
<10 GB est.	Non-sensitive data will be accessible to the public during and after the project. Sensitive data will be shared in compliance with the principles described in the CA.	Digital media of the WP3 Leader; folder FLEXX and TEAMS-MS
Name of the dataset	**Need and gaps in due diligence**	
Description	The WP3 dataset will contain chemical, biological and ecological data generated on each CS. These data are expected to be useful to partners involved in the WPs 1 and 2.	
Media Type	Texts, spreadsheets, presentations, images, graphs, infographics	
Language(s)	English	
Use & re-use	Assessment of needs and gaps in due diligence	
Size	<10 GB (est.)	
Format/license	.docx, pdf, .pptx, .xls., .jpg; png; etc.	
Version number	Not relevant	

Dataset reference and name: EDAPHOS – WP4 Organisation in charge: GIG		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Reports, prototypes, methods, spreadsheets with cross-sectional and/or panel data and algorithms, etc.	.docx, pdf, .pptx, .xls., .jpg; .png; etc.	MS Office (.docx, .pdf, .pptx,), Adobe Acrobat (.pdf reader or/and editor), picture manager to be defined for static graphics
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
<10 GB est.	Non-sensitive data will be accessible to the public during and after the project.	digital media of the WP4 Leader; folder FLEXX and TEAMS-MS

Funded by the European Union

	Sensitive data will be shared in compliance with the principles described in the CA.	
Name of the dataset	Need and gaps in due diligence	
Description	WP4 dataset will be a homogenized structure of environmental, economic and social data and variables at all CSs ensuring comparability for the econometric analysis of NBS in WP5.	
Media Type	Texts, spreadsheets, presentations, images, graphs, infographics	
Language(s)	English	
Use & re-use	Assessment of needs and gaps in due diligence	
Size	<1 GB	
Format/license	.docx, pdf, .pptx, .xls., .jpg; png; etc.	
Version number	Not relevant	

Dataset reference and name: EDAPHOS – WP5 Organisation in charge: EVO		
Data Type	Data Standards - Formats	Data Generation Software
• Reports, prototypes, methods, digital tools, videos. • Spreadsheets with time-series, cross-sectional data or/and panel data. • Data with special 1st-level classification, such as 5-stage life-cycle data that are further structured at 2nd-level in spreadsheet form.	.docx, pdf, .pptx, .xls., .csv any static visual picture able to process (e.g. .ai) any dynamic visual (e.g. .mp4, .mov, .wmv)	MS Office (.docx, .pdf, .pptx,), Adobe Acrobat (.pdf reader or/and editor), picture manager (e.g. ACD) to be defined for static graphics (.ai, .png, .jpeg), video manager (e.g. Windows Media Player) to be defined for movies (e.g. .mp4, .mov, .wmv)
Estimated Data Size	Data Sharing	Storage and Preservation
<10 GB	Data without confidentiality indication will be publicly accessible. Data indicated as confidential will follow the general process of the Consortium Agreement on IP and will be made available upon request and evaluation.	Mainly on EDAPHOS FLEXX and TEAMS-MS, as well as on any other offline or cloud storage of equivalent information security agreed by the consortium partners.

Name of the dataset	Need and gaps in due diligence
Description	• Internal due diligence: EDAPHOS datasets will be subject to internal peer-review and due diligence to ensure reliability of results, especially for those addressing to stakeholders and post-EDAPHOS commercial exploitation (e.g. ESS Treasuries, NBS Green Bonds). • External due diligence: Identified stakeholders and investors in the EDAPHOS proposal that will be participating as early adopters (e.g. EU Commission, EUSO, financial sector, municipal authorities) for which the data quality will have a direct impact will evaluate the data under confidentiality agreements.
Media Type	Texts, spreadsheets, presentations, images, videos, infographics
Language(s)	English
Use & re-use	Project development purposes, post-EDAPHOS exploitation plan upon bilateral or multilateral agreements between partners with consideration of IP protection.
Size	<10 GB (est.)
Format/license	.docx, pdf, .pptx, .xls., .csv any static visual picture able to process (e.g. .ai) any dynamic visual (e.g. .mp4, .mov, .wmv)
Version number	To be defined when data are generated

Dataset reference and name: EDAPHOS – WP6 Organisation in charge: LGI		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Analytical data, reports, documents, presentations	.docx, .xlsx, .pdf, .ppt, .png	Analytical data, reports, documents, presentations
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
1 GB max	Non-sensitive data will be accessible to the public during and after the project. However, sensitive data will be made available upon request as described in the CA.	WP6 folder FLEXX and TEAMS-MS
Name of the dataset	**Need and gaps in due diligence**	
Description	• The EDAPHOS C&D dataset will contain the data about C&D actions, supply chain traceability and scientific results compiled throughout the project. These data are expected to be useful to partners and WP leaders for the assessment of WP6 progress. • Multiple audiences beyond the project's own community, incl. media and the broad public, also those that may take an interest in the (potential) use of the results (e.g. the scientific community, industrial partner, policymakers) or interest in different material origins.	

Funded by the European Union

	• EDAPHOS website for public reports and OA online repositories such as Zenodo to make scientific publications accessible.
Media Type	Video, image, text, numerical data, n-grams, infographics
Language(s)	English
Use & re-use	C&D uses
Size	1 GB max
Format/license	.docx, .xlsx, .pdf, .ppt
Version number	Not relevant

Dataset reference and name: EDAPHOS – WP7 Organisation in charge: UMLP		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Reports, spreadsheets	docx, xlsx, pdf	MS Office
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
>1GB	Data will only be shared with Consortium	Project management folders and WP7 folder FLEXX and -MS
Name of the dataset	**Need and gaps in due diligence**	
Description	Spreadsheets that will help track contacts withing the consortium and monitor the status of deliverables and milestones within the project.	
Media Type	Spreadsheets	
Language(s)	English	
Use & re-use	Facilitate the communication between members and monitoring of the project	
Size	Less than 1GB	
Format/license	.xls, .txt	
Version number	Not relevant	

10 Annex III – Datasets that could be made openly accessible

This table below will be updated during the project.

Data producer	Brief description of dataset	Foreseen use & re-use	Possibility to share the data beyond the consortium
CSIC	The dataset EDAPHOS_DS_CSIC_20251002 corresponds to the data presented in the article with DOI: 10.3389/sjss.2025.14555	Data integration in metadatasets.	yes through ZENODO

11 Annex IV - Management of sensitive data

These guidelines are aligned to the European Union regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95.46.EC (General Data Protection Regulation). Referred to as from now onward as GDPR.

11.1 Definitions

"Personal Data"
Personal Data means any information relating to an identified or identifiable natural person ("**Data Subject**"), an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

"Data types used in the EDAPHOS project" shall be confined to the following main categories of Data Subject:
The names, e-mail and mail addresses, telephone numbers, job titles, job profiles, employer, classification of the knowledge system (type of research institution, type of hybrid knowledge institution), classification of the economic system (type of industry sector, type of industry umbrella organisation, type of commerce, type of technical sector), classification of the political system (type of executive system, type of legislative system), classification of civil society (type of civil society engagement, type of civil society funding institutions, type of community, individuals), classification of hidden actors (type of informal actors, type of criminal actors), geographical information.

"Controller, Data Controller, Processor, Data Processor, Data Subject" as set out in the relevant Data Protection Laws in force at the time.

"Data Discloser"
Data Discloser means a Party transferring the Personal Data to the Data Receiver. Any of the Parties may be a Data Discloser.

Funded by the European Union

"Data Receiver"
Data Receiver means a Party receiving the Personal Data from the Data Discloser. Any of the Parties may be a Data Receiver.

"Permitted Recipients"
Permitted Recipients means the Parties to this Agreement, the directors and employees of each Party, and any third parties that might be engaged to perform obligations in connection with the Project.

"Process/Processing"
Process/Processing means any operation or set of operations which is performed on Personal Data or on sets of Personal Data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

"Profiling"*
Profiling means any form of automated processing of Personal Data consisting of the use of Personal Data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour or movements.

11.2 Guidelines

Each partner is responsible for the Data Subjects with whom the individual partner enters into relationship, including the responsibility:

- to inform the „Data Subject" of the Processing of Personal Data and the rights of the Data Subject;
- to respond to Data Subjects when contacted by them in accordance with the GDPR;
- to ensure that the necessary preconditions exist for the Processing of Personal Data, including the obtaining of consent; and
- that data are erased when they are no longer necessary.

The nature of the relationship to the Data Subjects is explained in more detail in Annex 1.

If a Data Subject contacts one of the Partners in relation to EDAPHOS and GDPR, that partner has to immediately forward the request to the other partners. The partners will assist each other in responding to requests and provide the necessary information on the measures taken in accordance with Articles 15 to 22 GDPR.

11.2.1 Information to be provided according to Articles 13 and 14 GDPR

The Parties will provide the Data Subject with all information required under Articles 13 and 14 GDPR by means of a data protection notice on the EDAPHOS project website.

The Data Subject will be informed on the subject of the EDAPHOS project and advised that a participation is voluntary and does not involve any risks, costs or similar unpleasant consequences for them. Participation in the project and associated data collection is based on the consent of the Data Subject.

11.2.2 Obtaining Personal Data from other sources

The partner who obtains specific data from sources other than the Data Subject itself is responsible for informing the Data Subject accordingly and requesting separate permission for data use.

11.2.3 Principles and authority to process data

The partner who obtains specific Personal Data is responsible for ensuring:

a) that there is a valid legal ground for Processing and for documenting this to both supervisory authorities and the Data Subject;
b) that the data are not further processed in a manner that is incompatible with the purposes for which they were originally collected.

All partners shall include the Processing operations in their records in accordance with Article 30 paragraph 1 GDPR (as far as applicable to the Parties).

If a data protection impact assessment pursuant to Article 35 GDPR is required, it will be carried out jointly by the partners.

12 Annex V – Datasets to remain confidential

This table below will be updated during the project.

Purpose of processing :

- * Dataset ready to be uploaded on ZENODO
- ** Dataset will be made publicly available through ZENODO once published.

Data producer	Brief description of dataset and means for processing	purpose	Format	Lead partner	Origin
	WP1				
ONERA	EDAPHOS_DS1_ONERA_T1.1_202509 Soil spectral signature Library for composite soils (all CS)	**	.xls	GIG	T1.1
ONERA	EDAPHOS_DS2_ONERA_T1.1_202509 Vegetation spectral signature Library for CS1, 2 and 6	**	.xls	GIG	T1.1
ONERA	EDAPHOS_DS3_ONERA_T1.1_CS1_202602 Species inventory for CS1	**	.shp	GIG	T1.2
UMLP	EDAPHOS_DS4_UMLP_T1.1_CS3_202602 Species inventory for CS2	**	.shp	GIG	T1.1
ONERA	EDAPHOS_DS5_ONERA_T1.1_CS3_202602 Species inventory for CS3	**	.shp	GIG	T1.1
GIG	EDAPHOS_DS6_GIG_T1.1_CS4_202602 Species inventory for CS4	**	.xls	GIG	T1.1
GIG	EDAPHOS_DS7_ GIG _T1.1_CS5_202602 Species inventory for CS5	**	.xls	GIG	T1.1
UMLP	EDAPHOS_DS8_ UMLP _T1.1_CS6_202602 Species inventory for CS6	**	.xls	GIG	T1.1
GIG	EDAPHOS_DS9_ GIG _T1.1_CS7_202602 Species inventory for CS7	**	.xls	GIG	T1.1
ONERA	EDAPHOS_DS10_ONERA_T1.2_CS3_202602 RGB image representation of the hyperspectral acquisition and species classification map	*	.tif	ONERA	T1.2

Funded by the European Union

ONERA	EDAPHOS_DS11_ONERA_T1.2_CS3_202602 Soil type map	**	.tif	ONERA	T1.2
UMLP	EDAPHOS_DS12_ UMLP _T1.2_CS1_202602 UAV born RGB image and NDVI map for CS1	**	.tif	ONERA	T1.2
UMLP	EDAPHOS_DS13_ UMLP_T1.2_CS2_202602 UAV born RGB image, NDVI map and vegetation type clustering for CS2	**	.tif	ONERA	T1.2
GIG	EDAPHOS_DS14_ GIG _T1.2_CS4_202602 UAV born RGB image, NDVI map for CS4	**	.tif	ONERA	T1.2
GIG	EDAPHOS_DS15_ GIG _T1.2_CS5_202602 UAV born RGB image, NDVI map for CS5	**	.tif	ONERA	T1.2
UMLP	EDAPHOS_DS16_ UMLP _T1.2_CS6_202602 UAV born RGB image, NDVI map for CS6	**	.tif	ONERA	T1.2
GIG	EDAPHOS_DS17_ GIG _T1.2_CS7_202602 UAV born RGB image, NDVI map for CS7	**	.tif	ONERA	T1.2
ONERA	EDAPHOS_DS18_ONERA_T1.3_CS2_202602 Land cover map for CS2	**	.tif	ONERA	T1.3
ONERA	EDAPHOS_DS19_ONERA_T1.3_CS3_202602 Land cover map for CS3	**	.tif	ONERA	T1.3
	WP2 - Task2.1				
UMPL	EDAPHOS_DS1_Task2.1_datacollection_INERIS_CS1_UMLP_20240718 soil and vegetation indicators at To	**	.xls	INERIS	T2.1
CRES	EDAPHOS_DS2_Task2.1_datacollection_INERIS_CS2_CRES _20240718 soil and vegetation indicators at To	**	.xls	INERIS	T2.1
CSIC	EDAPHOS_DS3_Task2.1_datacollection_INERIS_CS3_CSIC _20240718 soil and vegetation indicators at To	**	.xls	INERIS	T2.1
GIG	EDAPHOS_DS4_Task2.1_datacollection_INERIS_CS4_GIG _20240718 soil and vegetation indicators at To	**	.xls	INERIS	T2.1
UNIBO	EDAPHOS_DS5_Task2.1_datacollection_INERIS_CS5_UNIBO _20240718 soil and vegetation indicators at To	**	.xls	INERIS	T2.1
UMLP	EDAPHOS_DS6_Task2.1_datacollection_INERIS_CS6_UMLP _20240718 soil and vegetation indicators at To	**	.xls	INERIS	T2.1
CRES	EDAPHOS_DS7_Task2.1_datacollection_INERIS_CS7_CRES _20240718 soil and vegetation indicators at To	**	.xls	INERIS	T2.1
	WP2 - Task2.2				
GIG	EDAPHOS_DS1_GIG_PIB_T2.2_CS1_20260216 ESS Data Repository for CS1	**	.xls	GIG	T2.2
GIG	EDAPHOS_DS2_GIG_PIB_T2.2_CS2_20260216 ESS Data Repository for CS2	**	.xls	GIG	T2.2
GIG	EDAPHOS_DS3_GIG_PIB_T2.2_CS3_20260216 ESS Data Repository for CS3	**	.xls	GIG	T2.2
GIG	EDAPHOS_DS4_GIG_PIB_T2.2_CS4_20260216 ESS Data Repository for CS4	**	.xls	GIG	T2.2
GIG	EDAPHOS_DS5_GIG_PIB_T2.2_CS5_20260216 ESS Data Repository for CS5	**	.xls	GIG	T2.2
GIG	EDAPHOS_DS6_GIG_PIB_T2.2_CS6_20260216 ESS Data Repository for CS6	**	.xls	GIG	T2.2
	WP3 - Task3.1				
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS1_WP3_T3.1_Soil_20240206 Soil parameters for poplar hybrid selection experiment at lab scale (all CS)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS2_WP3_T3.1_Plant_20240206 Plant parameters for poplar hybrid selection experiment at lab scale (all CS)	**	.xls	UMLP	T3.1

UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.1_Microbial Inoculation_Soil_20250703 Soil parameters for microbial inoculation experiment at lab scale (CS1, CS6)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.1_Microbial Inoculation_Plant_20250703 Plant parameters for microbial inoculation experiment at lab scale (CS1, CS6)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.1_Amendment Selection_Soil_202409 Soil parameters for amendment selection experiment at lab scale (CS1, CS6)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.1_Amendment Selection_Solution_202409 Soil solution parameters for amendment selection experiment at lab scale (CS1, CS6)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.1_Mustard Amendment test_Soil_202502 Soil parameters for Mustard Amendment test at lab scale (CS1, CS6)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.1_Mustard Amendment test_Plant_202502 Plant parameters for Mustard Amendment test at lab scale (CS1, CS6)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.1_Sulfur Amendment test_Soil_202507 Soil parameters for Mustard Amendment test at lab scale (CS1, CS6)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.1_Sulfur Amendment test_Plant_202507 Plant parameters for Sulfur Amendment test at lab scale (CS1, CS6)	**	.xls	UMLP	T3.1
UMLP, CRES, CSIC, UNIBO, GIG	EDAPHOS_DS3_WP3_T3.2_Pre-field experiment_Soil_202409 Soil parameters for pre-field experiment (All CS)	**	.xls	UMLP	T3.2
	WP4 and WP5				
UMLP	EDAPHOS_DS1_GIG_EVO_WP4_5_CS1_20251008 List of long-term variables	**	.xls	GIG EVO	WP4-5
CRES	EDAPHOS_DS2_GIG_EVO_WP4_5_CS2_20251008 List of long-term variables	**	.xls	GIG EVO	WP4-5
CSIC	EDAPHOS_DS3_GIG_EVO_WP4_5_CS3_20251008 List of long-term variables	**	.xls	GIG EVO	WP4-5
GIG	EDAPHOS_DS4_GIG_EVO_WP4_5_CS4_20251008 List of long-term variables	**	.xls	GIG EVO	WP4-5
UMLP	EDAPHOS_DS6_GIG_EVO_WP4_5_CS6_20251008 List of long-term variables	**	.xls	GIG EVO	WP4-5
CRES	EDAPHOS_DS7_GIG_EVO_WP4_5_CS7_20251008 List of long-term variables	**	.xls	GIG EVO	WP4-5

Funded by the European Union