



Call: **H2020-ICT-2018-20 / H2020-ICT-2020-2**

Grant Agreement no. **101016726**

**DELIVERABLE 5.2**  
**Data Management Plan**

**Start date of the project:** 1<sup>st</sup> January 2021

**Duration:** 48 months

**Project Coordinator:** Cédric ALLIER - CEA

**Contact:** Cédric ALLIER – [cedric.allier@cea.fr](mailto:cedric.allier@cea.fr)



## Document Classification

---

<b>Deliverable Number &amp; Title</b>	D5.2 Data Management Plan
<b>Work Package</b>	WP5 – Dissemination and Exploitation
<b>Author(s)</b>	ENS Lyon – Kiran PADMANABHAN
<b>Delivery due date</b>	2021-30-06
<b>Type</b>	Report
<b>Dissemination Level</b>	Public

## Document History

---

<b>Internal review</b>	2021-07-22	P1 CEA – Cédric ALLIER
<b>Contributions</b>	-	-
<b>Validation</b>	2021-07-22	Project Coordinator P1 CEA – Cédric ALLIER <a href="mailto:cedric.allier@cea.fr">cedric.allier@cea.fr</a>

## Document abstract

---

This deliverable presents the first version of the REVEAL Data Management Plan (DMP). It describes the guiding principle for data management and gives an overview of what data will be gathered and processed in the project, accordingly to the H2020 FAIR Data Management principle making data findable, accessible, interoperable and reusable.

The purpose of the DMP is to contribute to good data handling through indicating what research data the project expects to generate and describe which parts of the data that can be shared with the public. Furthermore, it gives instructions on naming conventions, metadata structure, storing of the research data and how to make public data available. This deliverable will be regularly updated during the project life.

The information contained in this report is subject to change without notice and should not be construed as a commitment by any members of the REVEAL Consortium. The REVEAL Consortium assumes no responsibility for the use or inability to use any procedure or protocol which might be described in this report. The information is provided without any warranty of any kind and the REVEAL Consortium expressly disclaims all implied warranties, including but not limited to the implied warranties of merchantability and fitness for a particular use.



**This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101016726**

# TABLE OF CONTENTS

1.	EXECUTIVE SUMMARY.....	7
2.	INTRODUCTION .....	7
2.1	WHY WOULD I WANT TO READ THIS DELIVERABLE? .....	7
2.2	INTENDED READERSHIP/USERS.....	7
2.3	OBJECTIVES AND SCOPE OF THE DOCUMENT .....	8
2.4	STRUCTURE OF THE DELIVERABLE.....	8
3.	DATA SUMMARY .....	8
3.1	DMP GUIDING PRINCIPLE.....	9
3.2	REVEAL DATA MANAGEMENT STRATEGY .....	9
3.3	REVEAL TYPE OF DATA .....	10
4.	FAIR DATA.....	11
4.1	MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA.....	11
4.2	MAKING DATA OPENLY ACCESSIBLE.....	12
4.3	MAKING DATA INTEROPERABLE.....	13
4.4	INCREASE DATA RE-USE (THROUGHT CLARIFYING LICENCES) .....	13
4.5	DMP REVIEW PROCESS & DATA INVENTORY .....	14
5.	ALLOCATION OF RESOURCES.....	15
6.	DATA SECURITY .....	15
7.	ETHICAL ASPECTS .....	16
8.	CONCLUSION .....	16
9.	REFERENCES .....	16
10.	ANNEX: DATASETS COLLECTED IN THE PROJECT.....	18

## List of Figures and Tables

---

Figure 1: open access to research data and publication diagram (source: Guidelines to the Rules on Open Access to Scientific publications and Open Access to Research Data in H2020) .....	9
Table 1: REVEAL research data.....	11

## List of Acronyms, Glossary

---

CA: Consortium Agreement

CC: Common Creative

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, Reusable

GA: Grant Agreement

GDPR: General Data Protection Regulation

IPR: Intellectual Property Rights

M: Month of the project

NDA: Non- Disclosure Agreement

ORDP: Open Research Data Pilot

PNB: Project NetBoard

V: Version

WP: Work Package

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). H2020 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. EXECUTIVE SUMMARY

The REVEAL Data Management Plan (DMP) follows the H2020 DMP template that was designed to be applied to any H2020 project that produces, collects or processes research data.

This DMP deliverable describes the data management principles and strategies, tools and REVEAL 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.

## 2. INTRODUCTION

### 2.1 WHY WOULD I WANT TO READ THIS DELIVERABLE?

---

It provides an easy overview of research data the 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.

### 2.2 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, in addition to Project handbook in deliverable 6.1 POPD Requirement no. 1 describing the project's ethics and privacy strategy, 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 REVEAL 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 store their data.



## 2.3 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 REVEAL project, as a H2020 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).

## 2.4 STRUCTURE OF THE DELIVERABLE

---

The document is structured following the guidelines of H2020 programme on FAIR Data Management in H2020 including the following information:

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

# 3. DATA SUMMARY

The REVEAL DMP aims to provide a strategy for managing key data generated and collected during the project and optimize access to and re-use of research data. The DMP is intended to be a 'living' document that will outline how the REVEAL research data will be handled during and after the project, and so it will be reviewed and updated at regular intervals.

All European Union funded projects must try to disseminate as much information as possible and on top of that the REVEAL project was signed up to the "Open Research Data Pilot (ORDP)" which means that we are committed to give open access to data generated during the project unless it goes against our legitimate interests. In this regard, the main purpose of the DMP is to ensure the accessibility and intelligibility of the data generated during the REVEAL project in order to comply with the Guidelines of the "Open Research Data Pilot". Each data set created during the project will be assessed and categorized as open, embargo or restricted by the owners of the content of the data set.

All the data sets, regardless of their categorization, will be stored in each of the participant entities databases and in the Project netboard (PNB) folder created as internal database of the partners. In addition, those categorized as open or embargo will be publicly shared (in the case of embargo, after the embargo period is over) through the public section of the project website and ZENODO (<https://zenodo.org/>).

ZENODO is an open access repository for all fields of science that allows uploading any kind of data file formats, which is recommended by the Open Access Infrastructure for Research in Europe (OpenAIRE).



## 3.1 DMP GUIDING PRINCIPLE

The DMP of REVEAL is realized within the Work Package 5.2.

The REVEAL project DMP follows the principle of Open Access according to the H2020 guidelines summarized in the diagram here below.

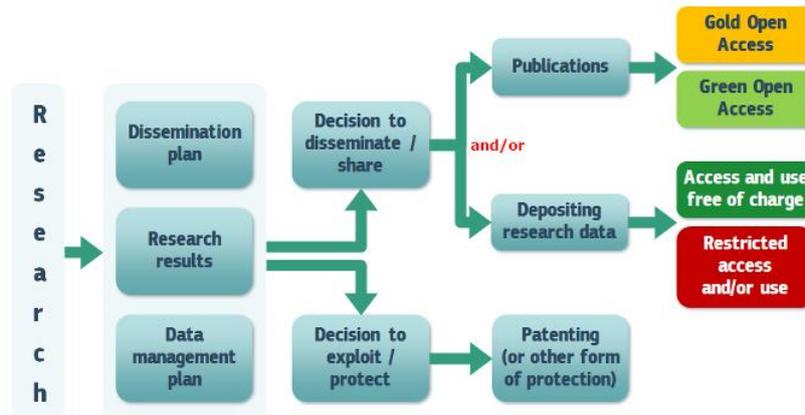


Figure 1: open access to research data and publication diagram (source: Guidelines to the Rules on Open Access to Scientific publications and Open Access to Research Data in H2020)

The other main principles for the REVEAL project DMP are the following:

- I. This DMP has been prepared by taking into account the template of the “Guidelines on Data Management in Horizon 2020”  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- II. The DMP is an official project Deliverable (D5.2) due at the end of month 6 (M6 June 2021) but it will be updated throughout the project. The first initial version will evolve depending on significant changes arising and periodic reviews at relevant reporting stages of the project.
- III. The consortium complies with the requirements of Regulation (EU) 2016/679 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 (General Data Protection Regulation - GDPR). Guidance on how these regulations interact with open-access data policy can be found here: <https://www.openaire.eu/ordp/>
- IV. Type of data, storage, confidentiality, ownership, management of intellectual property and access: procedures that will be implemented for data collection, storage, access, sharing policies, protection, retention and destruction will be in line with EU standards as described in the Grant Agreement (GA) and the Consortium Agreement (CA).

## 3.2 REVEAL DATA MANAGEMENT STRATEGY

As a project participating in the ORDP in H2020, the DMP’s Data Management strategy of REVEAL project is focused on the observation of FAIR (Findable, Accessible, Interoperable and Reusable) Data Management Protocols.

This document addresses for each data set collected, processed and/or generated in the project the following elements:

**Dataset reference and name:** Internal project Identifier for the data set to be produced. This will follow the format:



**ProjectName\_TaskNumber\_PartnerName\_DataSubset\_DatasetName\_Version\_DateOfStorage**, where the project name is REVEAL, the Partner Name represents the name of the data custodian (WP Leader / TASK Leader). (ex. [REVEAL]\_TASK5.1\_\_Dissemination activities)

**Dataset description:** description of the data generated or collected, including its origin (in cases where data is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the potential for integration and reuse.

**Standards and metadata:** reference to existing suitable standards. If these do not exist, an outline on how and what metadata will be created.

**Data sharing:** description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling reuse, and definition of whether access will be open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating the type of repository (institutional, standard repository for the discipline, etc.). In cases where the dataset cannot be shared, the reasons for this will be stated (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

**Archiving and preservation** (including storage and backup): description of the procedures to be put in place for long-term preservation of the data, including an indication of how long the data should be preserved, the approximate end volume, associated costs, and how these are planned to be covered.

### 3.3 REVEAL TYPE OF DATA

---

Annex I provides a list of all datasets currently expected to be generated in the REVEAL project and their planned accessibility. We recognise that this list will develop and grow as the project evolves.

Among the project datasets and deliverables, the following categories of outputs are declared “ORDP” that will be made “Open Access” (to be provided free of charge for public sharing).

These will be included in the ORDP and thus be managed according to the present DMP:

Project deliverable D5.1 Project website and social network account;

- Articles published in Open Access scientific journal;
- Conference and Workshop abstracts/articles.

Once generated (or collected), these data will be stored in several formats, which are: Documents, Images, Data, and Numerical codes.

In particular the following project deliverables are relevant:

#### D5.1: Website and logo

This deliverable is part of WP5 which is the main interface between the project and the general public, academics, and industrials. It describes the website dedicated to the REVEAL project, its logo and the key communication tools.

#### D5.3: Dissemination and exploitation

D5.3 aims at creating and publishing the public dissemination material (website, newsletters, etc.), adapting the dissemination support to the target, keeping track of publications and public disclosures. Management of Intellectual Property (IP) and patent submissions is also part of this deliverable.

#### D4.6: Integrative analysis of omics and imaging

The data related to this deliverable is a public report integrating imaging with omics analysis. In essence, this data validates the function of the microscope.



## D4.2: Murine models, cell lines and derived cells

The data is a public report (pdf) describing the mouse models and the cell lines derived from these mice.

Summarizing, REVEAL generates and collects the following research data relevant for the DMP:

Table 1: REVEAL research data

Title	WP No.	Lead Beneficiary	Nature
D5.1 - Website and logo	5	-----	Report
D5.3 - Dissemination and exploitation	5	-----	Report
D4.6 – Integrative analysis of omics and imaging	4	-----	Report
D4.2 – Murine models, cell lines and derived cells	4	-----	Report

### Upload instructions - Zenodo

Scientific publications, public deliverables and public datasets must also be uploaded to the [H2020 REVEAL Community](#) **AND** the [European Commission Funded Research \(OpenAIRE\) Community](#) in Zenodo.

To do this you must complete the following steps:

- Create a profile in Zenodo to be able to upload files
- <https://zenodo.org/communities/reveal-microscope/>
- Click on the REVEAL link above, or search for "H2020 [REVEAL]" under the "Communities" tab at the top of the Zenodo site
- On the Community site, click the green "New upload" button in the top right corner
- Enter requested data and confirm the upload. The information requested is located in the metadata list on SharePoint (*REVEAL Research Data*)
- Remember to add the European Commission community in the box labelled "communities". You can use the search function to locate the community and add it. The data will then automatically be uploaded to both communities, so you don't have to do it twice.

Uploading should be done as soon as possible and at the latest on article publication. Data Controllers are responsible for uploading datasets generated by them.

## 4. FAIR DATA

### 4.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR 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 (type of data, location, etc.) will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data.

The bibliographic metadata include all of the following:

- the terms "European Union (EU)" and "Horizon 2020";
- the name of the action, acronym and grant number;
- the publication date, and length of embargo period if applicable;
- a persistent identifier.

REVEAL open data will be collected in an open online research data repository: ZENODO.

<https://zenodo.org/communities/reveal-microscope/>

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 data sets, 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;
- 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 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:

- Neuronal microscopy
- Lensfree imaging
- Phase imaging
- Deep learning
- Convolutional neural networks
- Machine learning
- Omics

#### Naming conventions

Files and folders at data repositories will be versioned and structured by using a name convention consisting as follow: *FileType\_FileVersion/Number\_[REVEAL]\_PartnerNumber-PartnerName\_YYMMDD.FileExtension (ex. D1.5\_[REVEAL]\_P1-----\_YYMMDD.doc)*

FileType are:

- D stands for Deliverable
- DS stands for DataSet
- F stands for File (generic, ex. images, table, document)

#### Version numbers

Individual file names will contain version numbers that will be incremented at each revision (Vzz). 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.

## **4.2 MAKING DATA OPENLY ACCESSIBLE**

---

In order to maximise the impact of REVEAL 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 REVEAL 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 REVEAL dataset deliverables are both public (data access policy unrestricted) and they will be accessible by:

- REVEAL project web site
- Partners database
- OpenAIRE
- ZENODO (<https://zenodo.org>) for ORDP data and datasets
- Open access journals

All data deposited on Zenodo <https://zenodo.org/communities/reveal-microscope/> are accessible without restriction for public. For other data, 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.

#### 4.3 MAKING DATA INTEROPERABLE

---

Partners will observe OpenAIRE guidelines for online interoperability, including OpenAIRE Guidelines for Literature Repositories, OpenAIRE Guidelines for Data Archives, OpenAIRE Guidelines for CRIS Managers based on CERIF-XML. These guidelines can be found at: <https://guidelines.openaire.eu/en/latest/>.

Partners will also ensure that REVEAL data observes FAIR data principles under H2020 open-access policy:

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-datamgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-datamgt_en.pdf)

In order to ensure the interoperability, all datasets will use the same standards for data and metadata capture/creation.

As the project progresses and data is identified and collected, further information on making data interoperable will be outlined in subsequent versions of the DMP. In specific, information on data and metadata vocabularies, standards or methodology to follow to facilitate interoperability and whether the project uses standard vocabulary for all data types present to allow interdisciplinary interoperability.

#### 4.4 INCREASE DATA RE-USE (THROUGH CLARIFYING LICENCES)

---

Creative Common (CC) Licensing will be used to protect the ownership of the datasets. Different types of licenses will be considered.

However, an embargo period (usually 12 months for targeted journals) may be applied if the data (or parts of data) are used in published articles in “Green” open access journals.

For datasets deposited on a public data repository (i.e. Zenodo) the access is unlimited.

Restrictions on re-use policy are applied for all protected data (see Figure 1: Open access to research data and publication decision diagram), whose re-use will be limited within the project partners.

Other restrictions could include:

- the “embargo” period imposed by journals publication policy (Green Open access);
- some or all of the following restrictions may be applied with Creative Commons licensing of the dataset (attribution, non-commercial, share-alike, etc.).

Internal process of Quality evaluation is activated throughout the entire project duration to assess both project data /products and project process (See the D6.1 - Project Handbook). An internal peer review is performed for the main project deliverables to guarantee the deliverable is developed with a high level of quality. Each WP leader has to submit all the produced documents to another partner assigned as internal reviewer to check for the quality of the documents produced.

### Longevity

For data published in scientific journals, the underlying data will be made available no later than by journal publication. The data will be linked to the publication. Data associated with public deliverables will be shared once the deliverable has been approved and accepted by the EC. For other public datasets not directly linked to a scientific publication or deliverable, such datasets will be made available upon assessment by the Data Controllers that it is ready for publishing, and in the final month of the project at the latest.

Open data can be reused in accordance with the Creative Commons licences. Data classified as confidential will as default not be reusable due to privacy concerns.

The public data will remain reusable via Zenodo for at least 20 years. This is currently the lifetime stated by the host laboratory CERN. In the event that Zenodo has to close their operations, they have provided a guarantee that they will migrate all content (including metadata) to other suitable repositories.

## **4.5 DMP 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 D6.1 - Project Handbook).

Results data will be also analysed and collected throughout the project entire duration. To this purpose the Dissemination and Communication Report (See the D5.3 - Plans for dissemination of the results and training activities) 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 I), WP leaders 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.

## 5. ALLOCATION OF RESOURCES

Costs related to open-access to research data in H2020 are eligible for reimbursement under the conditions defined in the H2020 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 online data base: free of charge
- Copyright licensing with Creative Commons: free of charge.

The project member of General Assembly are also responsible of the Data Management of REVEALdataset and research data in accordance with each organization internal Data Protection Officer (DPO).

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).

The overall responsibility for data management lies with WP5 leader, Kiran Padmanabhan from ENS de Lyon.

## 6. DATA SECURITY

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 Project netboard (PNB) folder 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, on PNB like the use of 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.

## 7. ETHICAL ASPECTS

The WP7 aims at ensuring that ethical requirements are met for all research undertaken in the project, including data management aspects, in compliance with H2020 ethical standards.

All partners will assure that the EU standards regarding ethics and data management are fulfilled in compliance with the ethical principles (see Article 34) and confidentiality (see Article 36 as set out in the GA).

In addition:

1. In accordance with the GDPR 2016/679, the data controllers and processors are fully accountable for the data processing operations.
2. Templates for informed consent forms and information sheet are also available. More details in relation to Ethics (and Security) in relation to Data Management can be found in Section 5 of the GA.

See the following deliverables for more details:

- D7.1 HCT - Requirement No. 1
- D7.2 HCT - Requirement No. 2
- D7.3 POPD - Requirement No. 3
- D7.4 POPD - Requirement No. 4
- D7.5 A - Requirement No. 5
- D7.6 EPQ - Requirement No. 6
- D7.7 EPQ - Requirement No. 7

## 8. CONCLUSION

This document describes the main principles and guidelines for the Data Management for the REVEAL 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 and the data the description of dataset and research data gradually generated and collected.

## 9. REFERENCES

European Commission. *H2020 Programme. Guidance for the classification of information in research projects.*

European Commission. *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020.*



European Commission. *Horizon 2020 Programme. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020.*

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.



## 10. ANNEX: DATASETS COLLECTED IN THE PROJECT

Data set reference and name: ALS-Jena Person in charge: Susanne Freyberg		
Data Type	Data Standards - Formats	Data Generation Software
Reports, documentation and presentations	.docx, .xlsx, .pdf, .pptx	Microsoft word, excel, powerpoint, adobe acrobat reader
Estimated Data Size	Data Sharing	Data Storage and Preservation
~2 TB is estimated to be generated by this partner	The generated data will be shared on demand only between the [REVEAL] partners. The partners will use the Project netboard platform for data sharing, reports and deliverables. Published data will be made openly accessible in OpenAIRE, Zenodo, and, if applicable, respective open access journal repositories.	The generated data will be stored only in digital format (no physical support) on local servers. Back-ups are generated regularly. The scientific data will be stored during the entire project and for five years upon completion of the project.

Data set reference and name: WUT-DMP Person in charge: Wojciech Krauze		
Data Type	Data Standards - Formats	Data Generation Software
Programming files	Different calculation formats	Matlab, Python
3D CAD models	.IPT, .IAM, .IDW, .DWG, .IPN, .STL	Autodesk Inventor
Reconstructions from digital holographic microscope	.BMP, .MAT	Proprietary software, Python, Matlab
Reconstructions from holographic tomography	.MAT	Proprietary software, Python, Matlab
Analysis, data processing files	Different data processing formats	Excel, Matlab, Python
Reports, documentation and presentations	.docx, .xlsx, .pdf, .pptx, .odt, .ods, .odp	Microsoft word, excel, powerpoint, adobe acrobat reader, LibreOffice, ONLYOFFICE
Estimated Data Size	Data Sharing	Data Storage and Preservation
~50 TB is estimated to be generated by this partner	The generated data will be shared on demand only between the [REVEAL] partners. The partners will use the Project netboard platform for data sharing, reports and deliverables. Published data will be made openly accessible in OpenAIRE, Zenodo, and, if applicable, respective open access journal repositories.	The generated data will be stored only in digital format (no physical support) on local servers. Back-ups are generated regularly. The scientific data will be stored during the entire project and for five years upon completion of the project.



<b>Data set reference and name: Policlinico Milano - IRCCS Ca' Granda</b>		
Person in charge: Alessandro Cherubini		
<b>Data Type</b>	<b>Data Standards - Formats</b>	<b>Data Generation Software</b>
Programming files	Different calculation formats	JMP, R, Excel
NGS ANALYSIS	.FASTQ, .SAM, .BED, .BAM, .VCF, .BIM, .FAM	SAMTOOLS, GATK, BWA-MEM, BEDTOOLS, BCFTOOLS, VCFTOOLS, FASTQC, STAR, RSEM
Analysis, data processing files	Different data processing formats	Excel, Matlab, Python, FSL, SPM, JMP, R-Studio, P-LINK
Reports, documentation and presentations	.docx, .xlsx, .pdf, .pptx	Microsoft Word, Excel, PowerPoint, Adobe Acrobat Reader
<b>Estimated Data Size</b>	<b>Data Sharing</b>	<b>Data Storage and Preservation</b>
<10 TB is estimated to be generated by this partner	<p>The generated data will be shared on demand only between the [REVEAL] partners.</p> <p>The partners will use the Project netboard platform for data sharing, reports and deliverables.</p> <p>Published data will be made openly accessible in OpenAIRE, Zenodo, and, if applicable, respective open access journal repositories.</p>	<p>The generated data will be stored only in digital format (no physical support) on local servers. Back-ups are generated regularly and will be stored on external Hard-Disk . The scientific data will be stored during the entire project and for five years upon completion of the project.</p>

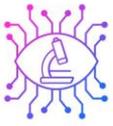
<b>Data set reference and name: CEA LETI REVEAL</b>		
Person in charge: Cédric ALLIER		
<b>Data Type</b>	<b>Data Standards - Formats</b>	<b>Data Generation Software</b>
Programming files	.txt .m .py	C++, Matlab, Python
Analysis, data processing files	.txt .mat .xlsx .png .tiff	C++, Matlab, Python
Reports, documentation and presentations	.docx, .xlsx, .pdf, .pptx .tex	Microsoft word, excel, powerpoint, adobe acrobat reader, Latex
<b>Estimated Data Size</b>	<b>Data Sharing</b>	<b>Data Storage and Preservation</b>
~100 TB is estimated to be generated by CEA LETI	<p>The generated data will be shared on demand only between the [REVEAL] partners.</p> <p>The partners will use the Project netboard platform for data sharing, reports and deliverables.</p> <p>Published data will be made openly accessible in OpenAIRE,</p>	<p>The generated data will be stored only in digital format (no physical support) on local servers. Back-ups are generated regularly. The scientific data will be stored during the entire project and for five years upon completion of the project.</p>



	Zenodo, and, if applicable, respective open access journal repositories.	
--	--	--

Data set reference and name: LMU Person in charge: Maria Robles		
Data Type	Data Standards - Formats	Data Generation Software
Orbitrap Exploris files	RAW	Excalibur
Proteomics data processing	MaxQuant output tables TXT ( <a href="http://www.coxdocs.org/doku.php?id=maxquant:table:directory">http://www.coxdocs.org/doku.php?id=maxquant:table:directory</a> )	Maxquant, Perseus, R, Python, Matlab
Reports, documentation and presentations	.docx, .txt, .xlsx, .pdf, .pptx	Microsoft word, excel, powerpoint, adobe acrobat reader
Estimated Data Size	Data Sharing	Data Storage and Preservation
~10 TB is estimated to be generated by this partner	<p>The proteomics raw and processed data will be available when needed to this grant REVEAL partners.</p> <p>In addition, proteomics raw and processed data will be deposited, and thus publicly available, in the ProteomeXchange Consortium via the PRIDE repository.</p> <p>Additional data and documents will be shared with the grant partners via the Project netboard.</p>	Raw and processed data will be stored, for a minimum of five years, on local and in long term at the Leibniz Supercomputing Center (LRZ) which provides secure back-ups.

Data set reference and name: ENS Lyon Person in charge: Kiran Padmanabhan		
Data Type	Data Standards - Formats	Data Generation Software
Programming files	Different calculation formats	Matlab, Python, R
Analysis, data processing files	.txt .mat .xlsx .png .tiff	C++, Matlab, Python
NGS ANALYSIS	.FASTQ, .SAM, .BED, .BAM, .VCF, .BIM, .FAM	SAMTOOLS, BEDTOOLS, BCFTOOLS, VCFTOOLS, FASTQC, STAR, RSEM
Analysis, data processing files	Different data processing formats	Excel, Matlab, Python, FSL, SPM, JMP, R-Studio,
Reports, documentation and presentations	.docx, .xlsx, .pdf, .pptx	Microsoft word, excel, powerpoint, adobe acrobat reader
Estimated Data Size	Data Sharing	Data Storage and Preservation
~50 TB is estimated to be generated by this partner	<p>The generated data will be shared on demand only between the [REVEAL] partners.</p> <p>The partners will use the Project netboard platform for data sharing, reports and deliverables.</p> <p>Published data will be made openly accessible in OpenAIRE, Zenodo, and, if</p>	The generated data will be stored only in digital format (no physical support) on local servers at the PSMN/IGFL/ENS Lyon. Back-ups are generated regularly. The scientific data will be stored during the entire project and for five



	applicable, respective open access journal repositories.	years upon completion of the project.
--	--	---------------------------------------