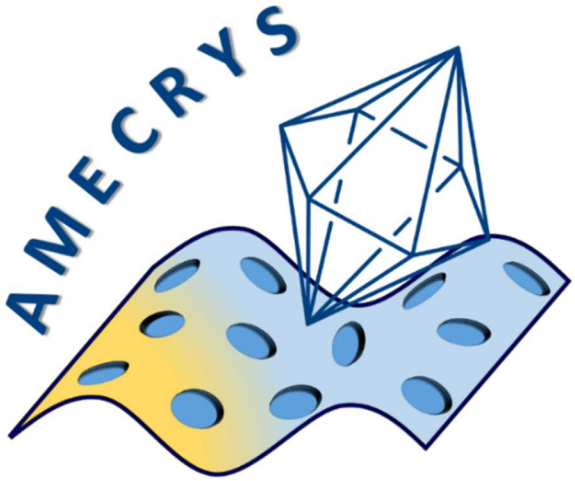
The AMECRYS project is funded by the European Union's Horizon 2020 programme under grant agreement No. 712965

Deliverable number	D7.2	Due date	March 31, 2017 (M6)
Deliverable title	Data Management Plan	Issue date	March 29, 2017
WP number	WP7	Author(s)	Joop ter Horst, Claire Lynch
Lead Beneficiary	UST	Reviewer(s)	All Partners
Deliverable type	ORDP	Status	Submitted
Dissemination level	Public		

AMECRYS - *Revolutionising Downstream Processing of Monoclonal Antibodies by Continuous Template-Assisted Membrane Crystallization*



Deliverables D7.2 - **Data Management Plan**

Version	Modifications	Date	Author(s)
1.0	Initial document creation	08/11/2016	Joop ter Horst (UST)
2.0	First draft	06/12/2016	Joop ter Horst, Claire Lynch (UST)
3.0	Final version	29/03/2017	Joop ter Horst, Claire Lynch (UST)

Table of Contents

1. Definitions

Dataset: Digital information created in the course of research but which is not a published research output. Research data excludes purely administrative records. The highest priority research data is that which underpins a research output. Research data do not include publications, articles, lectures or presentations.

Data Management Plan (DMP): A formal working document which outlines how datasets will be handled both during the active research phase and after the project is completed. DMPs in some form are now a requirement of a research grant proposals and therefore must be addressed at the earliest phase of the research lifecycle.

Digital Object Identifier (DOI): A persistent identifier for a document that can be handled by a resolution service to direct communications to the correct server. Developed by the International DOI Foundation (www.doi.org).

Metadata: Information about datasets stored in a repository/database template. For example, an image may require metadata that describe how large the picture is, the colour depth, the image resolution, when the image was created, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.

Repository: A digital repository is a mechanism for managing and storing digital content. Repositories can be subject or institutional in their focus (http://www.rsp.ac.uk/start/before-youstart/what-is-a-repository/).

Secondary data: Sources that contain commentary on or a discussion about a primary source.

Zenodo: a research output publication and archival service built and developed by researchers from the CERN IT department, to ensure that everyone can join in Open Science. Based on the open source Invenio digital library framework, Zenodo was created through the OpenAIRE project and is hosted at CERN (https://giving.web.cern.ch/content/zenodo-0).

A glossary of Open Access Abbreviations, Acronyms and Terms is available from SHERPA: Securing a Hybrid Environment for Research Preservation and Access (http://www.sherpa.ac.uk/glossary.html#d).

2. Introduction

This report describes the **Initial Data Management Plan (DMP)** for the AMECRYS project, funded by the EU's Horizon 2020 Programme under Grant Agreement number 712965. The purpose of the DMP is to set out the main elements of the AMECRYS consortium data management policy for the datasets generated by the project. The DMP will present in detail only the procedure for the management of datasets created during the lifetime of the project and describes the key data management principles, notably in terms of data standards and metadata, sharing, archiving and preservation.

This is the third version of the AMECRYS DMP and fulfils the project deliverable D7.2 month 6 by responsible partner University of Strathclyde (UST). It draws on Horizon 2020 guidance (European Commission, 2016) and guidance of the Research and Knowledge Exchange Services department, University of Strathclyde (University of Strathclyde (2014)).

Section 4 below list the project partners and key contacts. Section 5 provides a summary of the datasets generated during the lifetime of the project, including the types and format, the expected size of the datasets and the data utility. The specific description of how AMECRYS will make this research data findable, accessible, interoperable and reusable (FAIR) is outlined in section 6. Sections 7 to 9 outline the policy in relation to data resources, security and ethics.

This is a live "active" document to be updated at regular intervals during the project https://strathcloud.sharefile.eu/f/focfc1a4-6516-4951-93f0-4e86dd241d17.

3. Project Participants

Full Name	Short Name	Contacts
Consiglio Nazionale delle Ricerche, Italy	CNR	Dr Gianluca DI PROFIO g.diprofio@itm.cnr.it
Imperial College of Science, Technology and Medicine, UK	IMP	Dr Jerry HENG jerry.heng@imperial.ac.uk
Università della Calabria, Italy	UCAL	Prof Efrem CURCIO e.curcio@unical.it
Centre national de la recherche scientifique, France	CNRS	Dr Jean-Baptiste SALMON jean-baptiste.salmon-exterieur@solvay.com
Université libre de Bruxelles, Belgium	ULB	Dr Jim LUTSKO jlutsko@ulb.ac.be
University of Strathclyde, UK	UST	Prof Joop TER HORST Joop.terhorst@strath.ac.uk
Centre for Process Innovation Limited, UK	CPI	Dr John LIDDELL john.liddell@uk-cpi.com
GVS S.p.a., Italy	GVS	Dr Soccorso GAETA sng@gvs.com
Fujifilm Diosynth Biotechnologies, UK	FDB	Dr James PULLEN james.pullen@fujifilm.com

Coordinator Contact		
Dr Gianluca DI PROFIO	E: g.diprofio@itm.cnr.it T: +39 0984 492010/492014	Consiglio Nazionale delle Ricerche (CNR), Instituto per la Tecnologia delle Membrane (ITM), Via P. Bucci Cubo 17/C, I-87036 Rende (CS), Italy

4. Data Summary

The main purpose of the data collection/generation of this project is to industrially enable the template-assisted membrane crystallization process through a thorough scientific understanding of the process.

AMECRYS will produce several datasets during the lifetime of the project. The data will be both quantitative and qualitative in nature and will be analysed from a range of methodological perspectives for project development and scientific purposes. These will be available in a variety of easily accessible formats, including Post Script (PDF, XPS), Excel (XLSX, CSV), Word (DOC, RTF), Power Point (PPT), image (JPEG, PNG, GIF, TIFF), Origin (OPJ), compressed formats (TAR.GZ, MTZ), Program database (PDB).

Within AMECRYS approximately 49 separate datasets will be created (see list in table below). They are listed under each of the work package deliverables taken from the GA Annex 1 – Description of Action. The datasets will have the same structure, in accordance with the guide of Horizon 2020 for the Data Management Plan. The expected size of the datasets produced will be between 5MB and 1GB.

Table 5.1 – Potential datasets

Data Type	Format	Volume	IPR Owner
Work Package 2 - D2.1: Report on preparation of nanotemplates for mAb crystallization (lead: IMP)			
Experimental data – Brunauer–Emmett–Teller (BET)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – scanning electron microscope (SEM)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – transmission electron microscopy (TEM)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – crystal shape/size measurement	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – X-ray diffraction (XRD)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – High performance liquid chromatography (HPLC)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – dynamic light scattering (DLS)	XLSX, JPEG, PDF	< 100 MB	IMP
Synthesis protocol	DOC, PDF, JPEG	< 100 MB	IMP
Lab notes	DOC, PDF	< 100 MB	IMP
Work Package 2 - D2.2: HEL4 domain fragment & Anti CD20 mAb process specification report (lead: FDB)			
Final report	DOC, PDF	< 100 MB	FDB
USP Protocols and technology transfer package for production of Anti-CD20	DOC, XLSX , PDF, PPT	< 200 MB	FDB
DSP Protocols and technology transfer package for production of Anti-CD20	DOC, XLSX, PDF, PPT	< 200 MB	FDB
USP Protocols and technology transfer package for production of HEL4	DOC, XLSX, PDF, PPT	< 200 MB	FDB
DSP Protocols and technology transfer package for production of HEL4	DOC, XLSX, PDF, PPT	< 200 MB	FDB
Work Package 2 - D2.3: Report on the optimised nanotemplates for selective mAb recognition & crystallization (lead: IMP)			

Experimental data – Brunauer–Emmett–Teller (BET)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – scanning electron microscope (SEM)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – transmission electron microscopy (TEM)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – crystal shape/size measurement	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – X-ray diffraction (XRD)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – High performance liquid chromatography (HPLC)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – dynamic light scattering (DLS)	XLSX, JPEG, PDF	< 100 MB	IMP
Synthesis protocol	DOC, PDF, JPEG	< 100 MB	IMP
Lab notes	DOC, PDF	< 100 MB	IMP
Work Package 3 - D3.1: Report on pilot lines to prepare membranes available and debugged (lead: GVS)			
Experimental results from pilot tests	PDF	< 100 MB	GVS
Pictures of pilot plant developed	PDF	< 100 MB	GVS
Work Package 3 - D3.2: Report on the development of membranes for heterogeneous mAbs nucleation (lead: CNR)			
Experimental results - Membranes development data v1	DOC, XLSX, OPJ, PPT, JPEG, TIFF	< 500 MB	CNR
Experimental results - Membranes development data v2	DOC, XLSX, OPJ, PPT, JPEG, TIFF	< 1 GB	CNR
Experimental results - Heterogeneous nucleation data v1	DOC, XLXS, OPJ, PPT, JPEG, TIFF	< 500 MB	CNR
Experimental results - Heterogeneous nucleation data v2	DOC, XLSX, OPJ, PPT, JPEG, TIFF	< 1 GB	CNR
Work Package 3 - D3.3: Report on membrane preparation scaleup (lead: GVS)			
Report about prototype of membrane	PDF	< 100 MB	GVS
Work Package 4 - D4.1: Robust microfabrication protocols to embed hydrophobic fluoropolymers membranes within microfluidic chips (lead: CNRS)			
Fabrication Protocols of Microfluidic devices	PDF, JPEG	< 10 MB	CNRS
Work Package 4 - D4.2: Multilevel microfluidic device for high throughput crystallization screening (“pharma-on-a-chip” concept) (lead: CNRS)			
Microfluidic screening of membrane crystallization	PDF, XLSX	< 10 MB	CNRS
Work Package 4 -D4.3: Report on selective nucleation and growth kinetics of mAbs from screening tests (lead: UST)			
Experimental results - Crystallization Kinetics data v1	XLSX, JPEG	55 MB	UST
Experimental results - Crystallization Kinetics data v2	XLSX, JPEG	< 100 MB	UST
Work Package 5 -D5.1: Simulation code for thermodynamics of course-grained model of mAbs in confined geometry (lead: ULB)			
Code written by partners in the project - ftDFT Code	TAR.GZ	< 10 MB	ULB
Work Package 5 - D5.2: Open-source MCFFS computational simulation packages (lead: ULB)			
Code written by partners in the project - KMC Code	TAR.GZ	< 10 MB	ULB
Work Package 5 - D5.3: Report on multiscale simulation of mAbs crystallization on membranes/nanotemplates (lead: UCAL)			

Report on simulation activities - Models/software	PDF, JPEG	< 100 MB	UCAL
Work Package 5 - D5.4: Report on structural/ morphological/ bioactivity properties of mAbs crystals produced by prototype operation crystallizers (lead: CNR)			
Results of structural analysis	MTZ, PDB	50 MB	CNR
Results of crystal morphology	XLSX, TIFF	50 MB	CNR
Bioactivity tests	XLSX, TIFF	10 MB	CNR
Regression models & multivariate analysis	XLSX, TIFF, DOC	5 MB	CNR
Work Package 6 - D6.1: Design of continuous flow template assisted membrane crystallizer prototype (lead: UCAL)			
Design of prototype – Design specifications & Design drawings	PDF, JPEG	< 100 MB	UCAL
Work Package 6 - D6.2: Installation and validation of continuous flow template assisted membrane crystallizer prototype (lead: UCAL)			
Installation of prototype - Construction schedules	PDF, XLSX, CSV	< 100 MB	UCAL
Validation of prototype – Lab notes	PDF	< 500 MB	UCAL
Work Package 6 - D6.3: Report on prototype's operation monitoring & compliance with QS/GMP/CQA regulations (lead: CPI)			
Experimental results (USP/DSP/Analytics) – Anti-CD20 prototype operation	DOC, XLSX, PPT, PDF, XPS, JPEG, TIFF	500 MB	CPI
Experimental results (USP/DSP/Analytics) – HEL4 prototype operation	DOC, XLSX, PPT, PDF, XPS, JPEG, TIFF	500 MB	CPI
Report summarising operation performance including CQA, comparison to conventional approaches and GMP compliance	DOC	10 MB	CPI
Work Package 6 - D6.4: Techno-economic comparison between conventional batch and innovative DSP (lead: FDB)			
Final report	DOC, PDF	< 100 MB	FDB
Output from BioSolve Simulation	BioSolve, XLSX	< 200 MB	FDB

Table 5.2 – Lead Partners for Work Packages

Lead partner	**Related WP(s)**
IMP	WP2: Production of mAb/domain & Nanotemplates synthesis
GVS	WP3: Membranes development
CNRS	WP4: Microfluidics for continuous mAbs crystallization
ULB	WP5: Multi-scale modelling & characterization of mAb crystals
CPI	WP6: Prototype design, construction & operation

5 FAIR (Findability, Accessibility, Interoperability, and Reusability) data

5.1 Making data findable, including provisions for metadata

A DOI will be assigned to datasets for effective and persistent citation when it is uploaded to the repository [Zenodo]. This DOI can be used in any relevant publications to direct readers to the underlying dataset.

Each dataset generated during the project will be recorded in an Excel spreadsheet with a standard format and allocated a dataset identifier, see tables 6.1.1 and 6.1.2 below. The spreadsheet will be hosted at UST [ShareFile]. This dataset information will be included in the metadata file (see Section 6.3 and Annex B) at the beginning of the documentation, and updated with each version.

AMECRYS naming convention for project datasets will comprise of the following:

1. A unique chronological number of the datasets in the project will be added.
2. The title of the dataset.
3. For each new version of a dataset it will be allocated with a version number which will be for example start at v1.0.
4. A prefix "AM" indicating an AMECRYS dataset.
5. A unique identification number linking with the dataset work package and deliverable/task e.g., "W4_D4.3".

   01_ Crystallization Kinetics_v1.0._AM_W4_D4.3.xlsx

Search keywords will be provided when the dataset is uploaded to Zenodo which will optimise possibilities for re-use. Zenodo follows the minimum Data Cite metadata standards.

The specific metadata contents, formats and volume are given in table 6.1.1 below and will be further defined in future versions of the DMP.

Table 6.1.1: AMECRYS Datasets fields

Dataset Identifier	The ID allocated using the naming convention outlined in section 6.1
Tile of Dataset	The title of the dataset which should be easily searchable and findable
Responsible Partner	Lead partners responsible for the creation of the dataset
Work Package	The associated work package this dataset originates
Dataset Description	A brief description of the dataset
Dataset Benefit	What are the benefits of the dataset
Dataset Dissemination	Where will the dataset be disseminated
Type Format	This could be DOC, XLSX, PDF, JPEG, TIFF, PPT etc. (see table 5.1)
Expected Size	The approximate size of the dataset
Source	How/why was the dataset generated
Repository	Expected repository to be submitted
DOI (if known)	The DOI can be entered once the dataset has been deposited in the repository
Date of Repository Submission	The date of submission to the repository can be added once it has been submitted
Keywords	The keywords associated with the dataset
Version Number	To keep track of changes to the datasets
Link to metadata file	

Table 6.1.2: AMEYCRES Completed Dataset example

Dataset Identifier	01_ Crystallization Kinetics_v1.0._AM_W4_D4.3.xlsx
Title of the dataset	Crystallization kinetics
Responsible Partner	UST
Work Package	WP4, Deliverable 4.3
Dataset Description	A dataset of protein nucleation rates data measured from probability distributions of induction times
Dataset Benefit	The developed method for collection, processing and analysis of lysozyme data will apply for measurement of nucleation kinetics of Anti-CD20 protein. Furthermore, this data will be used to construct new heterogeneous protein nucleation theories (Task 5.2) and to enable process scale up (Task 6.1)
Dataset Dissemination	This data will be the basis for a peer reviewed journal paper on the heterogeneous crystal nucleation of the studied proteins
Type Format	Word, Excel, JPEG
Expected Size	< 100 MB
Source	Experimental results
Repository	Zenodo
DOI (if known)	To be inserted once the dataset is uploaded to the repository
Date of Repository Submission	In this example the expected date is 30-9-2020 or before if used for publication
Keywords	Protein Crystallization, Crystal Nucleation Rate, Induction Time Probability Distributions, Template Crystallization, Heterogeneous Nucleation
Version Number	V1.0
Link to metadata file	

5.2 Making data openly accessible

Research data which is created in the project is owned by the partner who generates it (GA Art. 26). Each partner must disseminate its results as soon as possible unless there is legitimate interest to protect the results. A partner that intends to disseminate its results must give advance notice to the other partners (at least 45 days) together with sufficient information on the results it will disseminate (GA Art. 29.1).

Research data should be deposited in the Zenodo repository as soon as possible unless a decision has been taken to protect results. Specifically, research data needed to validate the results in the scientific publications should be deposited in the data repository at the same time as publication (GA Art. 29.3).

During embargo periods, information about the restricted data will be published in the data repository, and details of when the data will become available will be included in the metadata. Where a restriction on open access to data is necessary, attempts will be made to make data available under controlled conditions to other individual researchers.

There will be three restricted datasets within deliverable D6.3 (lead: CPI):

- Anti-CD20 CHO cell line
- HEL4 E. coli strain
- Growth protocols (although basic data such as final product titre would be available).

These datasets are proprietary to Fujifilm (FDB) and may only be used in the restricted application of making material to support the work of this project. As these activities are enabling aspects of the project allowing the production of industry relevant biomolecules which can be used to develop the continuous crystallisation technology, it is not felt that restrictions will impact on eventual dissemination of the project outputs for the continuous crystallisation technology. CPI will archive the data as set out in Section 8: Data Security.

In accordance with GA Art. 25, data must be made available to partners upon request, including in the context of checks, reviews, audits or investigations. Data will be made accessible and available for re-use and secondary analysis.

The AMECRYS project has chosen to use Zenodo.org as the repository for storing the project data:

- *Research. Shared*. — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- *Citeable. Discoverable*. — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- *Communities* — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- *Funding* — identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- *Flexible licensing* — because not everything is under Creative Commons.

- *Safe* — your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

Not all project partners have access to an institutional repository and the use of Zenodo ensures data management procedures are unified across the project. A project page (community) has been setup for easy upload of project datasets https://zenodo.org/communities/amecrys-project-eu. Details of how to access the data will be available on the project web-site www.amecrys-project.eu. Zenodo.org is open, free, searchable and structured with flexible licensing allowing for storing all types of data: datasets, images, presentations, publications and software. In addition, Zenodo allows researchers to deposit both publications and data, while providing tools to link them.

All the public data of the project will be openly accessible at the repository. Non-public data will be archived at the repository using the “closed access” option.

Data objects will be deposited in ZENODO under:
- Open access to data files and metadata and data files provided over standard protocols such as HTTP and OAI-PMH.
- Use and reuse of data permitted.
- Privacy of its users protected.

Since the data is being deposited in an external repository [Zenodo], a dataset registry record should also be created in local host institutions repositories e.g. PURE for UST. The registry record should include relevant metadata explaining what data exists, and a DOI linking to where the data is available in the external repository. Any data which is deposited externally in a closed state, i.e. it is not accessible, should also be deposited in Pure, so that the University is still able to access the data. Here's an example of a dataset registry record in Pure, which includes a description of the dataset, and a DOI linking to where the data is available in the UK Data Service: https://pure.strath.ac.uk/portal/en/datasets/humour-styles-and-bullying-in-schools(fff279ab-3b66-4e25-99f1-39122b58839c).html

5.3 Making data interoperable

The AMECRYS project aims to collect and document the data in a standardised way to ensure that, the datasets can be understood, interpreted and shared in isolation alongside accompanying metadata and documentation.

Generated data will be preserved on institutional intranet platforms until the end of the Project (see section 8).

A metadata file will be created and linked within each dataset. It will include the following information:

General Information

- Title of the dataset
- Dataset Identifier
- Responsible Partner
- Author Information
- Date of data collection
- Geographic location of data collection
- The title of project and Funding sources that supported the collection of the data

Sharing/Access Information

- Licenses/access restrictions placed on the data
- Link to data Repository
- Links to other publicly accessible locations of the data
- Links to publications that cite or use the data
- Was data derived from another source?

Dataset/File Overview

- This dataset contains X sub-dataset as listed below
- What is the status of the documented data? – “complete”, “in progress”, or “planned”
- Are there plans to update the data?

Methodological Information

- Used materials
- Description of methods used for experimental design and data collection: <Include links or references to publications or other documentation containing experimental design or protocols used in data collection>
- Methods for processing the data: <describe how the submitted data were generated from the raw or collected data>
- Instruments and software used in data collection and processing-specific information needed to interpret the data
- Standards and calibration information, if appropriate
- Environmental/experimental conditions
- Describe any quality-assurance procedures performed on the data
- Dataset benefits

An example of a metadata file can be found in Annex B.

5.4 Increase data re-use (through clarifying licences)

The datasets will be made available for re-use through uploads to the Zenodo community page for the project.

In principle, the data will be stored in Zenodo after the conclusion of the Project without additional cost. All the research data will be of the highest quality, have long-term validity

and will be well documented in order other researchers to be able to get access and understand them after 5 years.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publically available data. Quality control of the data is the responsibility of the relevant responsible partner generating the data.

6. Allocation of resources

There are no immediate costs anticipated to make the datasets produced FAIR. The datasets will be deposited in the Zenodo repository for at least 5 years after the conclusion of the project. Any unforeseen costs related to open access to research data in Horizon 2020 are eligible for reimbursement during the duration of the project under the conditions defined in the Grant Agreement, in particular Article 6 and Article 6.2.D.3.

Prof Joop ter Horst and Claire Lynch based at the University of Strathclyde (UST) are responsible for data management within the AMECRYS project, specifically for D7.2 creation of data management plan and D7.6 updating the data management plan and ensuring the datasets are recorded. The PI of each partner will have overall responsibility for implementing the data management plan.

Each AMECRYS partner should respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with European Commission and local legislation. Dataset validation and registration of metadata and backing up data for sharing through repositories is the responsibility of the partner that generates the data in the WP.

The datasets in Zenodo will be preserved in line with the European Commission Data Deposit Policy. The data will be preserved indefinitely (minimum of 5 years) and there are currently no costs for archiving data in this repository.

7. Data security

For the duration of the project, datasets will be stored on the responsible partner's centrally provided storage, detailed in the table below.

Table 8: Data Storage

Short Name	Data Storage
CNR	All data is stored at the same time in internal servers of CNR-ITM, CNR-IC and CNR-IAC located in Rende and Bari, 300 km far each other. Data would also be fully copied in cloud-based repositories once provided by the central ICT system of CNR. Selected data is also stored in cloud-based repositories (Dropbox, Google drive) for sharing easily.
IMP	Non-sensitive data is stored in the "Home directory" of Imperial College London. Data stored on H: drives is secure and backed up daily, so a deleted file can be restored within 24 hours. More information can be found at https://www.imperial.ac.uk/admin-services/ict/self-service/connect-communicate/file-storage/home-directory-h-drive/. Sensitive data can be encrypted and then stored in the "Home directory" of Imperial College London. More information can be found at https://www.imperial.ac.uk/admin-services/ict/self-service/be-secure/protect-college-personal-information/encryption/
UCAL	At University of Calabria, data storage is managed by the ICT Center ("Centro ICT di Ateneo", c.ict@unical.it). The ICT Center is in charge of Security Analysis of Applications (AsIA), Virtual Desktop Infrastructure Service (Cloud-VDIS), S.O. Microsoft Windows/ S.O. Linux Debian Servers, High Performance Computing (HPC), Institutional Research Information System (Cineca IRIS, an innovative best-of-breed solution has been designed to fulfil the needs of academic and research institutions: IRIS is an IT platform that makes it easy to collect and manage data on research activities and outputs within an organization, https://www.cineca.it/en/content/iris-institutional-research-information-system).
CNRS	All data are either stored on Solvay's data repositories which are managed by the IT service of Solvay (with no access from outside) or on cloud-based repositories such as MyCore or Google drive for sharing easily the data.
ULB	All code and resultant data are stored in git repositories that are fully copied across numerous computers both on site at the ULB and off site. Selected data is also stored in cloud-based repositories (Dropbox, Google drive, ...).
UST	Data stored on the University of Strathclyde's storage is dual sited and replicated between two data centres which are physically separated by several hundred metres. Data links between datacentres are provided by dual disparate fabrics, providing added resilience. Additionally, the central I.T. service provides tape based backup to a third and fourth site. Data security is provided by access controls defined at a user level. The data will be stored on network drives http://www.strath.ac.uk/it/filestore/
CPI	The company has an IT group who have responsibility for IT infrastructure and data security. Electronic data is stored locally on network drives and/or data base systems (IDBS). Data is backed up daily to tape and stored in fireproof safes.
GVS	GVS has an internal archive where all proprietary data are stored. Moreover, all data are stored in secondary secure archives that are backed up every night.
FDB	Project data is stored on the internal intranet servers. FDB uses Commvault software to back up files on a nightly basis. This is a cloud-based solution and data is backed up to a data centre in London. Data on this server is also covered with our data recovery procedure which is replicated real time to the same data centre to enable remote access should we lose the server room.

Following completion of the project, all the responsibilities concerning data recovery and secure storage will go to the repository storing the dataset. Data will be archived and preserved in the Zenodo data sharing repository. This provides options for making data openly available and other data restricted access as required.

8. Ethical aspects

AMECRYS partners are to comply with the ethical principles as set out in Article 34 of the Grant Agreement, which states that all activities must be carried out in compliance with:

a. ethical principles (including the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity (European Science Foundation, 2011) — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct) and
b. applicable international, EU and national law.

The AMECRYS project does not involve the use of human participants or personal data in the research and therefore there is no requirement for ethical review.

Confidentiality

AMECRYS partners must retain any data, documents or other material as confidential during the implementation for the project. Further details on confidentiality can be found in Article 36 of the Grant Agreement along with the obligation to protect results in Article 27.

9. Other issues

As well as European Commission policies on open data management, Project Partners must also adhere to their own institutional policies and procedures for data management:

IMP
https://www.imperial.ac.uk/admin-services/ict/self-service/be-secure/protect-college-personal-information/sensitive-info/recommended-options/
https://www.imperial.ac.uk/admin-services/ict/self-service/be-secure/protect-college-personal-information/encryption/

UCAL
Source: http://www.unical.it/portale/ateneo/stat_reg/

- Regolamento per la gestione dell'innovazione e della proprietà intellettuale e industriale. Rectoral Decree n.1597, 19/10/2015
- Codice di comportamento dell'Università della Calabria. Rectoral Decree n. 2653, 23/12/2014

UST
http://www.strath.ac.uk/staff/policies/informationsecurity/
http://www.strath.ac.uk/media/ps/cs/gmap/academicaffairs/policies/research_code-of_practice_-_May_2010.pdf
http://www.strath.ac.uk/media/ps/cs/gmap/academicaffairs/policies/Research_Data_Policy_v1.pdf
http://www.strath.ac.uk/media/ps/cs/gmap/academicaffairs/policies/Research_Data_Policy_for_website.pdf

CPI
IT policies for the company are set out in written policies which are subject to periodic review

FDB
FDB has its own set of internal policies and procedures on data management.

10. References

European Commission (2016): Guidelines on Data Management in Horizon 2020, Version 2.1, 15 February 2016:
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Science Foundation (2011), European Code of Conduct for Research Integrity of ALLEA (All European Academies) and ESF, March 2011:
http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf

FAIR data principles (FORCE11 discussion forum):
https://www.force11.org/group/fairgroup/fairprinciples

OpenAIRE repository:
https://www.openaire.eu/opendatapilot-dmp

UK Data Service:
https://www.ukdataservice.ac.uk/manage-data/document

University of Strathclyde (2014) Research Data Deposit Policy, University of Strathclyde, Glasgow:
http://www.strath.ac.uk/media/ps/cs/gmap/academicaffairs/policies/Research_Data_Policy_for_website.pdf

Zenodo:
https://zenodo.org

11. Annex A - AMECRYS Grant Agreement extract

ARTICLE 27 — PROTECTION OF RESULTS — VISIBILITY OF EU FUNDING

27.1 Obligation to protect the results
Each beneficiary must examine the possibility of protecting its results and must adequately protect them — for an appropriate period and with appropriate territorial coverage — if:
a) the results can reasonably be expected to be commercially or industrially exploited and
b) protecting them is possible, reasonable and justified (given the circumstances).

When deciding on protection, the beneficiary must consider its own legitimate interests and the legitimate interests (especially commercial) of the other beneficiaries.

27.2 Agency ownership, to protect the results
If a beneficiary intends not to protect its results, to stop protecting them or not seek an extension of protection, the Agency may — under certain conditions (see Article 26.4) — assume ownership to ensure their (continued) protection.

27.3 Information on EU funding
Applications for protection of results (including patent applications) filed by or on behalf of a beneficiary must — unless the Agency requests or agrees otherwise or unless it is impossible — include the following:

"The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 712965".

27.4 Consequences of non-compliance
If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43). Such a breach may also lead to any of the other measures described in Chapter 6.

ARTICLE 29 — DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EU FUNDING

29.1 Obligation to disseminate results
Unless it goes against their legitimate interests, each beneficiary must — as soon as possible — 'disseminate' its results by disclosing them to the public by appropriate means (other than those resulting from protecting or exploiting the results), including in scientific publications (in any medium).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

A beneficiary that intends to disseminate its results must give advance notice to the other beneficiaries of — unless agreed otherwise — at least 45 days, together with sufficient information on the results it will disseminate.

Any other beneficiary may object within — unless agreed otherwise — 30 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed. In such cases, the dissemination may not take place unless appropriate steps are taken to safeguard these legitimate interests.

If a beneficiary intends not to protect its results, it may — under certain conditions (see Article 26.4.1) — need to formally notify the Agency before dissemination takes place.

29.2 Open access to scientific publications
Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results.
In particular, it must:
a) as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications; Moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications.
b) ensure open access to the deposited publication — via the repository — at the latest:

(i) on publication, if an electronic version is available for free via the publisher, or
(ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.

c) ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication.
The bibliographic metadata must be in a standard format and must 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, and
- a persistent identifier.

29.3 Open access to research data
Regarding the digital research data generated in the action (‘**data**’), the beneficiaries must:

a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:
   i. the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
   ii. other data, including associated metadata, as specified and within the deadlines laid down in the '**data management plan**';
b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.

See model grant agreement http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf

12. Annex B – Example Metadata File Template

This metadata file was generated on <insert date> by <insert name>

GENERAL INFORMATION

---

1. Title of Dataset:

2. Dataset Identifier in Repository:

2. Responsible Partner:

3. Author Information:

Investigator Contact Information
Name:
Email:

Supervisor Contact Information
Name:
Email:

Co-Supervisor Contact Information
Name:
Email:

3. Date of data collection:

4. Geographic location of data collection (where was data collected?):

6. The title of project and Funding sources that supported the collection of the data:

SHARING/ACCESS INFORMATION

---

1. Licenses/access restrictions placed on the data:

2. Link to data Repository:

3. Links to other publicly accessible locations of the data:

4. Links to publications that cite or use the data:

5. Was data derived from another source?
If yes, list source(s):

DATASET & FILE OVERVIEW

---

1. This dataset contains X sub-dataset as listed below:

A. Datasheet name:

B. Datasheet name:

C. Datasheet name:

D. Datasheet name:

2. What is the status of the documented data? – "complete", "in progress", or "planned"
Are there plans to update the data?

METHODOLOGICAL INFORMATION
---
1. Used Materials:

2. Description of methods used for experimental design and data collection:
<Include links or references to publications or other documentation containing experimental design or protocols used in data collection>

3. Methods for processing the data: <describe how the submitted data were generated from the raw or collected data>

4. Instruments and software used in data collection and processing-specific information needed to interpret the data:

5. Standards and calibration information, if appropriate:

6. Environmental/experimental conditions:

7. Describe any quality-assurance procedures performed on the data:

8. Dataset Benefit: