ICM - Institut du Cerveau et de la Moelle épinière: DMP ICM

1- INFORMATION ABOUT THE PROJECT

Project acronym

Name of the research project

Identifier of the call(s) for proposal and project funder(s)

Recommandations:
Please specify :
the call for proposal id and the project funder(s) id.

Project's Coordinator information

Recommandations:
Please specify :

- *Full name*
- *Email*

Organization and unit of the coordinator

Recommandations:
please detail the:
Organization / Unit or lab / City / Country:

ICM partner(s) if different from coordinator

Recommandations:
1- Indicate only those involved in data collection, analysis or processing. Name and unit.
2- indicate: Organisation / Unit or lab / City / Country

Project dates and duration

Scope of the project

- Local
- National
- EU
- International
- Monocentric
- Multicentric
- Multi partners

Recommandations:
Example: local monocentric / national multicentric / European / International (if extra European: specify the countries)

2- MANAGEMENT PLAN AT ICM

Identification and affiliation of the author of the DMP

Recommandations:
Please specify : Team, Core facility or Department

Date of creation of DMP (1st version)

Current version

Recommandations:
Please specify :

- *Version n°:*

- *Date:*

Reference to national/funder/sectorial/departmental procedures for data management

Recommandations:

Please specify : Standards used

Indicate the data management standards that you will follow in case they are public.

3- DATA COLLECTION

Purpose of the data collection/generation in relation with the objectives of the project

Origin of the data in case of new collection

Status of the data in case of existing dataset

Recommandations:

Please specify :

- *Status*
- *Origin*
- *Publications DOI (if applicable)*
- *Terms of re-use*

If the dataset relies on the re-use of existing data, indicate their origin and the publications related to the dataset (DOI). If material protected by specific rights is used during the project, specify the access or re-use restrictions.

Explain, if needed, the terms of re-use and the means used to ensure compliance with those terms.

In case you plan to add new data to these existing datasets, state if the licence allows it

Type and format of the data collection

Recommandations:

Please specify :

- *Type of data*
- *Format of data*

Data collection period

Exemple de réponse:

Example: 2019-2020

Recommandations:

Please indicate the start and finish dates.

Dataset production methods

Recommandations:

Please indicate the methods and protocols for data collection, acquisition, transformation

Example: Illumina HiSeq 2000 Sequencing System, MRI....

Expected size/volume of the data

- <50MB
- 50-100 MB
- 100 MB-1G
- 1G-10 GB
- 10-50 GB
- >50 GB

Data utility and to whom

Recommandations:

Describe which audience might potentially be interested in the produced dataset, in addition to the partners of the project. This is to appreciate how the public dissemination of data, at the end of the project, is envisaged

Data raising specific ethical issues and personal data protection

- Personal data
- Animal data
- Omics data

Recommandations:

- Personal data: nominative, anonymized / aggregated

- Data collected during animal testing validated by an ethics committee Research with partners outside Europe : recommendations of the H2020 manual have been respected

- Omics data from genetic resources obtained in the context of the Nagoya protocol are subject to the same rules, particularly with regard to access and benefit sharing (ABS). It must be ensured, before the end of the project, that the partners have access to the digital resources they are entitled to according to the BFF consortium agreement, and know how to use the associated analysis tools (assess the need for documentation and training).

Please refer to EU H2020 guide: How to complete your ethics self-assessment (02/201)

4- DATA MANAGEMENT DURING PROJECT

Who is responsible for data management during the project?

Recommandations:

repeat as many as implicated teams

Exemple de réponse:

- X is responsible for data collection, processing and analysis

-Y is responsible for data storage, archiving and sharing

What hardware or software resources do you need to store and manage your data?

Exemple de réponse:

- Data will be stored on the internal storage area, with user access control and sharing permissions. A procedure for directories and files naming, and rules for metadata documentation, will be shared within projects members.

- microscopy images will be stored and managed on OMERO, on a dedicated and domain specific server, implemented and maintained by the IT department. OMERO manages groups and user access permissions, import, export, visualisation, annotations, metadata standard...

- neuroimaging data will be stored on XNAT, a dedicated and domain specific tool, implemented and maintained by the ITdepartment. XNAT manages groups and users access permissions, import, export, visulaisation, annotations, metadata standard...

- clinical data will be captured and stored using Redcap, a Research Electronic Data Capture, which is implemented and supported by the IT department. Groups and users access permissions, data management and quality checks are featured in Redcap.

Recommandations:

- data and metadata must be stored on central and secure space. always ask your IT department for more details. if you use standard management tools like: OMERO, XNAT or REDCAP data and metadata will be securely stored.

- Be careful, you absolutely must not store your data on internet storage space (dropbox, google drive, ...)!

Do you use any naming convention and versioning system for your data?

Exemple de réponse:

Each file is named as follows: project_taxonomy_date_version

Versions are identified as: V01, V02, DV (draft version, FV (final version)

Recommandations:

BP_nommage_pantheonsorbonne

https://library.stanford.edu/research/data-management-services/data-best-practices

Which metadata standards do you use? If you don't use metadata standard, outline what type(s) of metadata will be created and how.

Do you provide a supplementary documentation in order to describe more precisely your data?

Exemple de réponse:

*ISA '**I**nvestigation', '**S**tudy' and '**A**ssay' model will be as metadata orgnisation.*

MINSEQE describes the Minimum Information about a high-throughput nucleotide SEQuencing Experiment

DICOM is the standard used for neurominaging data. The XNAT tool manages the dicom format, and allows exports in BIDS struture, easy for data publication.

OME-TIFF standard and CMPO ontology will be used to describe microscopy image metadata and resulting phenotypes. OMERO server uses the OME-XML and OME-Tiff and allows additionnal tags and annotaions.

Recommandations:

Metadata are data used to define or describe other data, regardless of the type of medium used (printed or electronic). The use of standard metadata is necessary to make your data visible, findable and reusable.

Metadata standards are templates that specify all the metadata required to describe a resource.

When describing your data, you should give preference to your discipline's standards, ontologies or vocabularies

Validated and recognised standards can be found here ontologies at https://www.ebi.ac.uk/ols/ontologies

What is the quality control procedure of the data?

Exemple de réponse:

A dedicated data manager will plan quality checks and data accuracy. RedCap offers a quality assurance workflow to help clinical data studies ensuring completness and correctness.

A standard operating procedure exists and will be followed by engineer responsible for data management.

Recommandations:

- Try to answer the questions : How do you make sure that the data in your research project are properly managed? How often do you check it and what do you do in case of discrepancy?

- Contact your Quality depatment.

5- DATA SHARING, DISSEMINATION AND RE-USE

Data sharing during the project

Exemple de réponse:

The whole dataset will be made available to all the partners of the project for consultation only. Each partner will remain the only responsible institution for managing the datasets collected at their level.

REDCAP allows group and members management, to securely access and share clinical data.

XNAT allows group and members management, to securely access and share neuro-imaging data and meta-data. It is not opened for external access, but VPN access can be guaranted by the IT department, to access icm servers in secure way.

OMERO allows group and members management, to securely access and share microscopy image data and meta-data. it is not opened to external access, but VPN access can be used to access icm servers.

Recommandations:

Data sharing and access arrangements for the project partners, or other third-parties, have to be defined in advance and fully described. Define clearly what is shared: raw data, aggregated data, results of analysis.

For data transfer, different solutions are available at ICM:

Globus to transfer high volume of data: transferred data are encrypted, access is define by the owner - Refer to Sharing files with Globus

SFTP solution - HOWTO – SFTP Connection

Data sharing, dissemination and re-use after the project (FAIR)

Recommandations:

-Datasets to be shared: data produced and/or used in the project which will be made openly available as the default.

- Recommendation: If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions. In multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

- Open access or restricted access: conditions for access / controlled access procedure details (i.e. a machine readable license) / existence of a data access committee / methods or software tools needed to access the data? (Generally, web access, or open source code)

- Useful resource for licence types used for dissemination: https://b2share.eudat.eu

Data reading:

Specify what documentation or software is needed to understand and access the data (codes, abbreviations, versions of the software for reading, explanatory documents, etc.). The documentation associated with a survey data includes the questionnaire and the pollster manual.

Recommendation:

EU H2020 programme: The Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020

Refer to point to CC-0 or CC-BY as a straightforward and effective way to make it possible for others to mine, exploit and reproduce the data

Sensitive data:

Existence of sensitive data justifying waiving of the dissemination principle: protected, personal, strategic data, or data from private partnerships, etc

Recommendation:

- check public repositories (by domain) recommended by the journal where the results will be published,

- check the availability and detailed information about the selected repository on FAIRsharing.org or https://www.re3data.org/

- We recommand the use of european repositoties when available.

- if no domain specific repository is known, use Zenodo at https://zenodo.org/

Exemple de réponse:

At the end of the project, raw data from RNA sequencing will be hosted at the The European Nucleotide Archive (ENA), under CC BY 4.0 licence. The repository uses a predefined and rich list of metadata.

Microscopy images will be hosted at the public repository[the Image Data Resource](#) (IDR).

6- ARCHIVING AND PRESERVATION

Storage and backup during the project: Describe the storage solution and the backup frequency

Exemple de réponse:

Personal computer of XXX / external drive of the XXX team / external drive synchronized with server XXX / storage space of the XXX data center....

Please specify :

- Data to be archived long term
- *Backup frequency*

Recommandations:

ICM offers a scientific storage capacity based on Apache Lustre files system solution. Lustre is an open-source, distributed parallel file system software platform designed for scalability, high-performance, and high-availability.

Access is allowed per team or per person

Choices at ICM for back-up:

ICM backup solution based on Open IO

Cloud public solution (AWS solution may be proposed)

At ICM, those needs have to be described in the ICM specifications form (refer to your Fiche d'Expression de Besoins - FEB)

Data destruction procedure: IT offers to delete data on demand through the ticketing system on due date (link to retention duration)

Long-term preservation after the project

Recommandations:

Recommendation: You should only archive data that might be reused in future research; are unique, non reproducible or difficult to reproduce; necessary for a global understanding of the research; necessary to validate the results presented in a scientific publication.

Archiving media: please specify if you use the method (local server, distant server, external drive, repository)

Recommended lifetime: taking into account the existing legal or prescribed requirements

Data repository: Name of the repository or cloud services where the data will be deposited, if it is identified. Specify the arrangements in place with the identified repository.

Comment: Long-term preservation is a task for the research institution. Alternatively, data collected in the framework of a limited (in scope and in time) project can be stored via the scientific journal storage facilities, if and when the results are published. Most journals have now specific requirements or recommendations. Some funding bodies, as well, provide lists of recommended repositories.

Useful resources:

[DataCite](#) - [Directory of research data repositories](#)

FAIRsharing.org - [Directory of life sciences repositories](#)

Institut Pasteur - [Comment choisir un entrepôt de données](#)

Recommendation: Preference should be given to certified repositories supporting open access where possible,

Data to be destroyed: Specify whether some types of data will be destroyed (e.g. personal data, according to CNIL recommendations, for example)

Recommendation: Anticipate the monetary, material and human costs associated with data archiving and preservation, and arrangements made to cover them (in particular after the end of the project).

UK Data Service – [Data Management Costing tool and checklist](#)

Useful resources:

Cines - [Preservation](#)

[European Open Science Cloud](#)

Exemple de réponse:

Data to be archived long term:

Archiving media:

Recommended duration:

Volume of archived data:

Data repository:

Data to be destroyed:

Budget allocated to the preservation: