Neurospin Horizon Europe DMP example

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Horizon Europe DMP (english)" fourni par Commission européenne.

Plan Details

Plan title	Neurospin Horizon Europe DMP example
Deliverable	0.1
Version	First version
Fields of science and technology (from OECD classification)	5.1 Psychology and cognitive sciences
Language	eng
Creation date	2024-01-17
Last modification date	2024-07-08
Identifier type	URL
License	**Name** Etalab Open License 2.0 **URL** http://spdx.org/licenses/etalab-2.0.json

Project Details

Project title	Neurospin Horizon Europe DMP example
Acronym	Neurospin_Horizon_Europe_DMP
Abstract	**[Here describe your own project]** The goal of this project is to provide a DMP example to start from, as a basis for your own Horizon Europe projects. The focus is on: • monocentric projects, • data acquired at Neurospin, • the kind of data we usually manipulate: MRI, MEG, EEG, demographics, omics, questionnaires, clinical data, ... Multicentric projects, involving data transfers between partners and multiple IT infrastructures, warrant a specific DMP.
Start date	2024-01-01
End date	2027-12-31

Research outputs :

1. Demographic information describing study participants (Dataset)

2. Functional and structural MRI images, using 3 T MRI (Image)
3. High-resolution functional and structural MRI images, using 7 T MRI (Image)
4. Magnetoencephalography data (Dataset)
5. Behavioural data (Dataset)
6. Stimulation software programmed in Python (Software)
7. Custom Python scripts running fMRIprep for fMRI preprocessing (Software)
8. Custom Python scripts using ScikitLearn and NiLearn for fMRI analysis (Software)

Contributors

Name	Affiliation	Roles
Papadopoulos Orfanos Dimitri - 0000-0002-1242-8990	NEUROSPIN - 201722552U	• DMP Manager • DMP Manager • Personne contact pour les données (Demographics, MRI 3T, MRI 7T, MEG, Behavioural, Stimulation, fMRI preprocessing, fMRI analysis)

Neurospin Horizon Europe DMP example

1. Data Summary

Demographic information describing study participants

- **Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.**
- **What types and formats of data will the project generate or re-use?**
- **What is the purpose of the data generation or re-use and its relation to the objectives of the project?**
- **What is the expected size of the data that you intend to generate or re-use?**
- **What is the origin/provenance of the data, either generated or re-used?**
- **To whom might your data be useful ('data utility'), outside your project?**

Will you reuse?

[The answer depends on your research question and the availability and quality of relevant datasets]

- We need to acquire new data because no available dataset addresses our research question or this specific pathology.
- We will use a subset of UKBiobank as control data. The subset has been obtained under authorisation [...] and is made available for 3 years with an estimated cost of [...].

What types of data?

Demographic data can be stored in tabular form, using the TSV file format.

Purpose of the data and its relation to the objectives of the project?

[That you have to write yourself! We recommend you pre-register your research and analysis plan and refer to it when explaining the purpose of the data and its relation to the objectives of the project.]

We use demographic data to characterise and stratify the participants in our study.

What is the expected size of the data?

Tabular data should hardly exceed 100 KB.

What is the origin/provenance of the data?

The sponsor of this longitudinal, monocentric study is CEA. We will recruit 100 volunteers of African descent, and attempt to achieve a 50:50 female/male balance and evenly distributed ages between 18 and 50 years old. We will collect their demographic data.

We will reuse UK Biobank demographic data.

To whom might your data be useful, outside your project?

Researchers interested in [the same scientific domain?] can reuse the data to address similar scientific questions.

Functional and structural MRI images, using 3 T MRI

- **Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.**
- **What types and formats of data will the project generate or re-use?**
- **What is the purpose of the data generation or re-use and its relation to the objectives of the project?**
- **What is the expected size of the data that you intend to generate or re-use?**
- **What is the origin/provenance of the data, either generated or re-used?**
- **To whom might your data be useful ('data utility'), outside your project?**

Will you reuse?

[The answer depends on your research question, and the availability and quality of relevant datasets]

- We need to acquire new data because no MRI datasets address our research question or this specific pathology.
- We will use a subset of UKBiobank as control data. The subset has been obtained under authorisation [...] and is made available for 3 years with an estimated cost of [...].

What types of data?

We will collect:

- MRI raw data:
  - raw data in DICOM format produced by our Siemens Prisma Fit machine
  - derived data in NIfTI format
- Behavioural data: software specific format (.xyz format of XYZ Inc.) and derived tabular data in TSV format
- Physiological data: software-specific and derived tabular data in TSV format

The MRI data are the bulk of the data we plan on collecting in this study.

Raw MRI data are archived as DICOM files internally. Because DICOM files are complex and may contain proprietary binary blobs, DICOM files are not shared and raw data are converted early on to the NIfTI file format, a domain standard, and organised according to the BIDS specification. MRI data are preprocessed, and the results are shared as "derivatives" alongside raw data, as suggested by the BIDS specification.

Purpose of the data and its relation to the objectives of the project?

[That you have to write yourself! We recommend you pre-register your research and analysis plan and refer to it when explaining the purpose of the data and its relation to the objectives of the project.]

Anatomical and functional MRI data are our main tool to explore structures and functions of the brain in vivo.

What is the expected size of the data?

[Your mileage may vary. Here, I have used mean sizes for datasets acquired in 2024 on the 3T Prisma Fit.]

The size of raw data is around 4 GB per participant, 2 GB for DICOM files and 1.5 GB for NIfTI files. As for derivatives, we expect they account for up to 10 GB per participant. However, the challenge in our domain is not the mere size of data in GB, but the myriad of small files produced during preprocessing.

What is the origin/provenance of the data?

The sponsor of this longitudinal, monocentric study is CEA. We will recruit 100 volunteers of African descent, and attempt to achieve a 50:50 female/male balance and evenly distributed ages between 18 and 50 years old. We will acquire data over two time points, *baseline* and *follow-up*, and shall strive for an interval of 8 weeks between these two time points.

To whom might your data be useful, outside your project?

Researchers interested in [the same scientific domain?] can reuse the data to address similar scientific questions.

High-resolution functional and structural MRI images, using 7 T MRI

- **Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.**
- **What types and formats of data will the project generate or re-use?**
- **What is the purpose of the data generation or re-use and its relation to the objectives of the project?**
- **What is the expected size of the data that you intend to generate or re-use?**
- **What is the origin/provenance of the data, either generated or re-used?**
- **To whom might your data be useful ('data utility'), outside your project?**

Will you reuse?

[The answer depends on your research question, and the availability and quality of relevant datasets]

- We need to acquire new data because no MRI datasets address our research question or this specific pathology.
- We will use a subset of UKBiobank as control data. The subset has been obtained under authorisation [...] and is made available for 3 years with an estimated cost of [...].

What types of data?

MRI data are the bulk of the data we plan on collecting in this study:

- Structural and functional MRI data:
  - The IDEA software provided by the data acquisition system manufacturer, Siemens Healthliners, reconstructs MRI images and stores them in proprietary format in a local database on the MRI consoles.
  - Raw data is sent on demand to the storage system of NeuroSpin, in DICOM format, for long-term archival.
  - Because DICOM files are complex and may contain proprietary binary blobs, DICOM files are converted early on to the NIfTI file format, a domain standard, and organised according to the BIDS specification. MRI data are preprocessed, and the results are shared as "derivatives" alongside raw data, as suggested by the BIDS specification.
- Detailed description of the MRI protocol, as proprietary Siemens EDX or EXAR files, and as PDF exports.
- Behavioural data: proprietary format (.xyz format of XYZ Inc.) and derived tabular data in TSV format.
- Physiological data: proprietary and derived tabular data in TSV format.

Purpose of the data and its relation to the objectives of the project?

[That you have to write yourself! We recommend you pre-register your research and analysis plan and refer to it when

explaining the purpose of the data and its relation to the objectives of the project.]

Structural and functional MRI data are our main tool to explore structures and functions of the brain in vivo.

What is the expected size of the data?

[Your mileage may vary. Here, I have used mean sizes for datasets acquired in 2024 on the 7T Investigational Device.]

The size of raw data is around 10 GB per participant, 6.5 GB for DICOM files and 3.5 GB for NIfTI files. As for derivatives, we expect they account for up to 20 GB per participant. However, the challenge in our domain is not the mere size of data in GB, but the myriad of small files produced during preprocessing.

What is the origin/provenance of the data?

The sponsor of this longitudinal, monocentric study is CEA. We will recruit 100 volunteers of African descent, and attempt to achieve a 50:50 female/male balance and evenly distributed ages between 18 and 50 years old. We will acquire data over two time points, *baseline* and *follow-up*, and shall strive for an interval of 8 weeks between these two time points.

We will reuse UK Biobank data.

To whom might your data be useful, outside your project?

Researchers interested in [the same scientific domain?] can reuse the data to address similar scientific questions.

Magnetoencephalography data

- **Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.**
- **What types and formats of data will the project generate or re-use?**
- **What is the purpose of the data generation or re-use and its relation to the objectives of the project?**
- **What is the expected size of the data that you intend to generate or re-use?**
- **What is the origin/provenance of the data, either generated or re-used?**
- **To whom might your data be useful ('data utility'), outside your project?**

Will you reuse?

[The answer depends on your research question and the availability and quality of relevant datasets]

- We need to acquire new data because no available dataset addresses our research question or this specific pathology.

What types of data?

We acquire and store MEG data as FIFF files. This format is widely used and understood by processing software.

Purpose of the data and its relation to the objectives of the project?

[That you have to write yourself! We recommend you pre-register your research and analysis plan and refer to it when explaining the purpose of the data and its relation to the objectives of the project.]

MEG data are our main tool to explore functions of the brain with high temporal resolution in vivo.

What is the expected size of the data?

[Your mileage may vary. Here, I have used a typical size of datasets acquired on the Elekta MEG.]

The size of raw data is around 5 GB per participant. As for derivatives, we expect they account for up to 10 GB per participant.

What is the origin/provenance of the data?

The sponsor of this longitudinal, monocentric study is CEA. We will recruit 100 volunteers of African descent, and attempt to achieve a 50:50 female/male balance and evenly distributed ages between 18 and 50 years old. We will acquire data from each participant within a single session of 2 hours.

To whom might your data be useful, outside your project?

Researchers interested in [the same scientific domain?] can reuse the data to address similar scientific questions.

Behavioural data

- **Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.**
- **What types and formats of data will the project generate or re-use?**
- **What is the purpose of the data generation or re-use and its relation to the objectives of the project?**
- **What is the expected size of the data that you intend to generate or re-use?**
- **What is the origin/provenance of the data, either generated or re-used?**
- **To whom might your data be useful ('data utility'), outside your project?**

Will you reuse?

[The answer depends on your research question and the availability and quality of relevant datasets]

- We need to acquire new data because no available dataset addresses our research question or this specific pathology.

What types of data?

Behavioural data can be stored in tabular form, using the TSV file format.

Purpose of the data and its relation to the objectives of the project?

[That you have to write yourself! We recommend you pre-register your research and analysis plan and refer to it when explaining the purpose of the data and its relation to the objectives of the project.]

We investigate individual behaviour.

What is the expected size of the data?

[Your mileage may vary. Measuring response times over multiple reptitions, as opposed to answering a question, might result in larger files.]

Tabular data should hardly exceed 1 MB.

What is the origin/provenance of the data?

The sponsor of this longitudinal, monocentric study is CEA. We will recruit 100 volunteers of African descent, and attempt to achieve a 50:50 female/male balance and evenly distributed ages between 18 and 50 years old. We will collect behavioural data.

To whom might your data be useful, outside your project?

Researchers interested in [the same scientific domain?] can reuse the data to address similar scientific questions.

Stimulation software programmed in Python

- **Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.**
- **What types and formats of data will the project generate or re-use?**
- **What is the purpose of the data generation or re-use and its relation to the objectives of the project?**
- **What is the expected size of the data that you intend to generate or re-use?**
- **What is the origin/provenance of the data, either generated or re-used?**
- **To whom might your data be useful ('data utility'), outside your project?**

Question sans réponse.

Custom Python scripts running fMRIprep for fMRI preprocessing

- **Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.**
- **What types and formats of data will the project generate or re-use?**
- **What is the purpose of the data generation or re-use and its relation to the objectives of the project?**
- **What is the expected size of the data that you intend to generate or re-use?**
- **What is the origin/provenance of the data, either generated or re-used?**
- **To whom might your data be useful ('data utility'), outside your project?**

Will you reuse?

We plan on reusing state-of-the-art processing pipelines:

- fMRIprep
- sMRIprep

What types of data?

The above pipelines will be run from scripts that will be published on GitHub.

We plan on packing the software and scripts in containers, to be able to rerun them in the future and reproduce the results. The recipes for the containers will be published as well, and the resulting containers made available alongside the data.

Purpose of the data and its relation to the objectives of the project?

[That you have to write yourself! We recommend you pre-register your research and analysis plan and refer to it when explaining the purpose of the software and its relation to the objectives of the project.]

Preprocess structural and functional MRI data.

What is the expected size of the data?

The size of software sources is negligible.

To whom might your data be useful, outside your project?

Anyone who wishes to reproduce or extend our results.

Custom Python scripts using ScikitLearn and NiLearn for fMRI analysis

- **Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.**
- **What types and formats of data will the project generate or re-use?**
- **What is the purpose of the data generation or re-use and its relation to the objectives of the project?**
- **What is the expected size of the data that you intend to generate or re-use?**
- **What is the origin/provenance of the data, either generated or re-used?**
- **To whom might your data be useful ('data utility'), outside your project?**

Will you reuse?

We plan on reusing state-of-the-art processing software:

- scikit-learn
- Nilearn

What types of data?

We will write scripts using the above libraries, and develop in-house software.

We plan on packing the software and scripts in containers, to be able to rerun them in the future and reproduce the results. The recipes for the containers will be published as well, and the resulting containers made available alongside the data.

Purpose of the data and its relation to the objectives of the project?

[That you have to write yourself! We recommend you pre-register your research and analysis plan and refer to it when explaining the purpose of the software and its relation to the objectives of the project.]

Process structural and functional MRI data.

What is the expected size of the data?

The size of software sources is negligible.

To whom might your data be useful, outside your project?

Anyone who wishes to reproduce or extend our results.

2. FAIR data

2.1. Making data findable, including provisions for metadata

- **Will data be identified by a persistent identifier?**
- **Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.**
- **Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?**
- **Will metadata be offered in such a way that it can be harvested and indexed?**

Persistent identifier

[Decide on the partitioning of the dataset: a single dataset/identifier of multiple subsets/identifiers per modality?]

A DOI will be assigned when the dataset is published.

Rich metadata

We will organise our data according to the BIDS specification, which allows for a JSON file describing the columns of tabular data such as TSV files.

Search keywords

We will use the very keywords associated with our scientific articles.

Metadata

The BIDS specification is a de facto standard in our scientific domain. We can extract relevant metadata and adapt it to the needs of the publishing platform we will use.

2.2.1. Making data accessible : Repository

- **Will the data be deposited in a trusted repository?**
- **Have you explored appropriate arrangements with the identified repository where your data will be deposited?**
- **Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?**

Trusted repository

Pseudonymised data will be made available for scientific research, within the boundaries set by the informed consent, and will initially remain on an internal storage system at CEA, dedicated to research data. We shall make data available once quality is judged satisfactory, after QA, preprocessing, and QC. This usually coincides with the publication of scientififc articles.

We plan on making our data accessible on EBRAINS, which:

- is a EU-funded research infrastructure for neuroscience;
- strives to be compatible with the GDPR;
- provides assistance to curate the data before storage and dissemination;
- provides DOIs to identify dataset.

Appropriate arrangements

We plan on following the process described in the Integrate and share your data section of EBRAINS.

2.2.2. Making data accessible : Data

- **Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.**
- **If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.**
- **Will the data be accessible through a free and standardized access protocol?**
- **If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?**
- **How will the identity of the person accessing the data be ascertained?**
- **Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?**

Openly available, embargo

Directly identifying personal data are used for research management purposes and cannot be shared. They are kept in a dedicated encrypted space on an internal CEA server, accessible only by a dedicated clinical team, for the sole purpose of recruiting participants and managing data acquisition.

Pseudonymised data will be made available for scientific research, within the boundaries set by the informed consent, and will initially remain on an internal storage system at CEA, dedicated to research data. We shall make data available once quality is judged satisfactory, after QA, preprocessing, and QC. This usually coincides with the first associated article published.

Free and standardized access protocol

[Two options here: send the data to the algorithms, and bring the algorithms to the data.]

We shall publish our data using the standard EBRAINS mechanisms - typically HTTPS access.

Restrictions on use, data access committee

Since the data we acquire can be viewed as personal health data, thus sensitive data, access will be restricted. Access to the data will be granted after examination of research projects by a scientific committee and an executive committee that will make sure data sharing remains within the bounds set by regulatory rules for personal data:

- reuse within the limits of the informed consent,
- signature of a Data Transfer Agreement by the recipients.

Identity of the person accessing the data

Data Transfer Agreements are signed by a director with the power of signature for any organisation that asks permission to access the data. Individual accounts are associated with a professional mail address associated with the organism, and that professional mail address is verified.

2.2.3. Making data accessible : Metadata

- **Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?**
- **How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?**
- **Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?**

Openly available

In France, data from the public sector are usually made available under the Etalab Open License, but we will examine the possibility of a CC0 license specifically for metadata. The metadata will include information on the process to follow to gain access to the data.

How long

By default, data can be kept for a maximum of 20 years after it has been published, and lust then be discarded.

Software

We use standard and documented file formats, as well as the BIDS specification for file organisation, that do not require a specific piece of software.

2.3. Making data interoperable

- **What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?**
- **In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?**
- **Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?**

What data and metadata vocabularies, standards, formats or methodologies

We will follow the BIDS specification for data and metadata organisation.

Qualified references to other data

We cannot redistribute UK Biobank data, but we shall clearly cross-reference that dataset and explain why and how we use it.

2.4. Increase data re-use

- **How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?**
- **Will your data be made freely available in the public domain to permit the widest re-use possible? Will**

your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?
- **Will the data produced in the project be useable by third parties, in particular after the end of the project?**
- **Will the provenance of the data be thoroughly documented using the appropriate standards?**
- **Describe all relevant data quality assurance processes.**
- **Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.**

Provide documentation

We shall either write a data article or add an annex documenting the data to the main scientific article.

The BIDS specification allows for detailed information such as definitions and units of measurement.

Freely available

Public domain is not an option, due to the sensitive nature of our data. However, we will publish the data for further reuse, within the limits set by personal data protection regulatory rules.

Usable by third parties

According to current regulatory rules, which restrict the reuse of personal data.

Provenance of the data

Scientific articles detail the general origin of the data. We can associate to the DOI all the persons implicated in the data acquisition process and their respective roles.

Data quality assurance process

A dedicated clinical team supervises the recruitment and reception of study participants, based on documented internal procedures.

Data acquisition is supervised by one of the scientists responsible for the study, and a radiographer in the case of MRI. A radiologist checks anatomical MRI images for abnormal, unexpected features.

Allocation of resources, data security and ethical aspects

For data acquired using the NeuroSpin platform, the calculated costs include the price of storage.

The IT infrastructure of CEA is implemented according to the information systems security policy (PSSI) of CEA.

All projects involving data acquisition are validated by a national or local ethics committee, according to national laws.

---

3. Other research outputs

- **In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).**
- **Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.**

Research outputs include the software used to preprocess and analyse our dataset. Where possible, we plan on reusing state-of-the-art software. The scripts that run existing pipelines, as well as the sources of software we develop ourselves will be versioned and made available on GitHub.

The main pieces of software are described as research output on their own.

---

4. Allocation of resources

- **What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?**
- **How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)**
- **Who will be responsible for data management in your project?**
- **How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs**

and potential value, who decides and how, what data will be kept and for how long)?

Costs

The NeuroSpin platform factors in the costs of internal data storage and usage when calculating its overall expenses. However, the costs associated with personal health data sharing and reuse are still under evaluation. Publishing such data presents a challenge due to the difficulty of planning and covering associated costs in an evolving regulatory landscape. We plan on relying on standardised platforms, backed by EU subsidies, such as EBRAINS, to mitigate these expenses.

Responsible for data management

?

Long term preservation

National regulations typically limit the default long-term preservation period for personal health data to 20 years. In addition to internal storage, we plan on relying on standardised platforms, backed by EU subsidies, such as EBRAINS, and we are open to utilizing other compliant data-sharing platforms that we expect to emerge at the national, European, or organizational level.

---

5. Data security

- **What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?**
- **Will the data be safely stored in trusted repositories for long term preservation and curation?**

Data security

The IT infrastructure of CEA adheres to the information systems security policy (PSSI) of CEA. At NeuroSpin, we store data on a departmental storage system dedicated to scientific data. We plan on evolutions to take into account more stringent personal health data protection rules. We perform daily backups stored in separate locations within the CEA Saclay research centre.

Trusted repositories for long term preservation and curation?

We recognize the importance of long-term data preservation and curation. Again, we plan on relying on standardised platforms, backed by EU subsiding, such as EBRAINS, and more generally on data sharing platforms that we expect to emerge at the national or European level, or at the level of our organisation.

---

6. Ethics

- **Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).**
- **Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?**

Ethics and legal issues that impact data sharing

We handle personal data, and singularly personal health data. Strict regulatory rules govern the sharing of such data.

Research protocols involving human subjects have a clearly identified principal investigator, validated by a national ethics committee (CPP) and authorised by the national public establishment responsible for the safety of health products (ANSM). These protocols comply with the MR-001 reference framework.

Informed consent for data sharing and long term preservation

Yes, our informed consent clearly explains data sharing, long-term preservation, and the possibility of data being made available outside the EU.

---

7. Other issues

- **Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?**