

Université Gustave Eiffel: Gustave Eiffel University - DMP template

1. PRESENTATION OF THE COLLECTED DATA SETS

Collected data sets

Recommandations:

Give a summary of the data that will be collected during the project

2. PLANNING DATA COLLECTION

Data responsibility

Recommandations:

In the vast majority of cases, data is not covered by intellectual property law. The institution that collects the data or mandates the data collection is responsible for the data. In a project, certain data sets may be under the responsibility of a single party and others under the responsibility of multiple. Data responsibility is generally specified in the consortium agreement. Explain who is responsible for the various data sets in the project.

Exemple de réponse:

- Université Gustave Eiffel
- Data is under the co-responsibility of Université Gustave Eiffel and the project partners. Article 7 of the project consortium agreement discusses the question of the ownership of the results.
- Data produced by XXX, available as open data under the Open Database Licence, will be reused for the project.
- Certain data belonging to XXX and shared with Université Gustave Eiffel under the terms of agreement no. XXX will be reused.

Mode of collection

Recommandations:

Data may be collected by sensors (specify what they are), surveys, or software (specify the software name and version). List here all the means for collecting data for the project.

Nature of data

Exemple de réponse:

2D/3D/HD mapping data used for algorithms to calculate the specific position

Recommandations:

It may be images, videos, maps, textual data, numbers, etc.

Content of the data set

Exemple de réponse:

The data collected for the project includes personal data. This data is recorded in the Université Gustave Eiffel data processing register. Register form n°XXX was completed and sent to the university's Data Protection Officer.

Recommandations:

In this part, specify if the data is protected by regulation (GDPR, copyright, confidentiality: medical, statistical, industrial, defence, legal)

Are you collecting personal data? Please complete the record sheet available here : <https://intranet.univ-eiffel.fr/universite/affaires-juridiques-et-institutionnelles/protection-des-donnees-personnelles-rgpd/fiche-dinscription-au-registre-des-activites-de-traitement-de-luniversite-gustave-eiffel>

For more information : <https://intranet.univ-eiffel.fr/universite/affaires-juridiques-et-institutionnelles/protection-des-donnees-personnelles-rgpd>

Data set format (if digital)

Recommandations:

Opt for open formats, so as to offer open data and/or reuse of the data. The [French General Interoperability Framework](#) and the [CINES](#) specify the standards in place.

In general, choose standard, open formats to facilitate long-term data preservation.

If the format is not standard and open, provide documentation describing the format specifications, or make the format self-documented (e.g. tags in language, use of metadata, etc.) to facilitate long-term preservation.

Exemple de réponse:

- TXT (ASCII)
- CSV
- HTML
- ODT
- TIFF
- PNG
- PDF
- XML

- The proprietary format of GNSS receiver data may require the use of converters, but certain open-source software programs can open certain proprietary formats (e.g. RTKLib, GoGPS) and allow proprietary formats to be converted to standard formats (RINEX, CSV).

Data format (if not digital)

Exemple de réponse:

Notes, documents, photos, film rolls, VHS

Language

Exemple de réponse:

English is used to document the data

3. DOCUMENTATION AND METADATA

Naming convention

Recommandations:

Using a naming convention guarantees that team members and data users can easily find the data file that they are looking for. In general, it is advised to include the date, the type of data or a brief description of the data, and the version number. If the naming rules are not easy to understand, explain the reasoning clearly.

Exemple de réponse:

Files are named in the following way:

<Sensor>_<DD_MM_YYYY>_<version-number>.format . DD_MM_YYYY corresponds to the last date that the file was updated

Metadata

Recommandations:

The metadata (data describing the data and data set) is also very useful, to make the data findable and interoperable. There are standard descriptors (like Dublin Core, MARC, TEI, etc.), either general or discipline-specific. To find more information about discipline-specific standards: <https://fairsharing.org/>, <https://www.dcc.ac.uk/guidance/standards/metadata> or <http://rd-alliance.github.io/metadata-directory/>.

Exemple de réponse:

- Surveys are carried out in line with the DDI standard.
- Data sets are described using Dublin Core metadata.

Documentation

Recommandations:

Documentation makes it easier to understand and eventually reuse the data. It may include reasons for collecting the data (objectives of the research project) and the means implemented (methodology, instruments, etc.).

A data set may be accompanied by documentation or specific instructions: e.g. documentation explaining the structure of files, units of stored values, descriptive elements of values (1 = present, 0 = absent), etc. The best practice is to use an open format (CSV, PDF, etc.).

Exemple de réponse:

Documentation included with survey data includes the questionnaire and survey manual.

4. STORING, USING AND PROCESSING DATA

Estimating the volume of the data set

Recommandations:

This estimation allows you to plan your storage needs.

Exemple de réponse:

- Approximately 250 GB
- 5 TB

Storing a data set (digital) - Storage infrastructure

Recommandations:

Storage of confidential, personal or sensitive data on Cloud platforms (e.g. Dropbox, Google Drive, etc.) is forbidden by the PSSIE, for security reasons. See the Security Policy for State Information Systems, conveyed by the Circular of the Prime Minister n° 5725/SG of 17 July 2014. <https://www.legifrance.gouv.fr/circulaire/id/38641>

Long-term storage of data on personal computers, USB sticks or external hard drives is not recommended, also for security and long-term access reasons.

Exemple de réponse:

- Raw data is stored on the hard drive of the collection tools.
- The processed data is stored on the laboratory's network: link to the network.

Storage of data (if non-digital) - Storage conditions

Recommendations:

Depending on the data medium (paper, etc.), describe the facility and how the data is stored.

Backing-up a data set (digital)

Recommendations:

A back-up solution on top of the original storage location may be useful, in the event of any incident.

Exemple de réponse:

- Two forms of storage are planned: 1) under the responsibility of the WP manager, who will ensure that the data is regularly backed-up, 2) on the omnispace platform dedicated to the project.
- External back-up hard drive stored in the data manager's office (XXX building 2nd floor)

Controlling access during the project

Exemple de réponse:

Access to data sets subject to industrial confidentiality will be limited. Only authorised people will have access rights for the network on which the data is stored.

Sharing during the project

Recommendations:

Sharing data while the project is underway should be done in line with the ownership agreements referenced in the first section. Sharing confidential, personal or sensitive data on Cloud platforms (e.g. Dropbox Drive, etc.) is forbidden by the PSSIE, for security reasons.

To find the most suitable solution (FTP network access, Eiffel Cloud, Filesender, USB stick or external hard drive, specific website, etc.), contact your IT technician or the Delegated General Directorate for Information and Digital Technology (DGDIN).

Exemple de réponse:

The entire project team can access the data set via the platform provided by the Institute for Applied Research and Experimentation in Civil Engineering (IREX).

Traceability and quality assurance

Recommendations:

Guaranteeing traceability and data quality is very important for scientific integrity (respecting the values of reliability, rigour, transparency and even honesty).

Exemple de réponse:

A lab book is regularly used to ensure that data processing methods are kept traceable. Data quality is guaranteed by various systems for prevention (metrological monitoring of instruments, knowledge of uncertainties, etc.) and checking (detection of abnormal or missing values, harmonisation of experimental protocols, sharing databases). For mass spectrometry data, quality and compliance will be checked by various quality control methods and regular calibration of instruments.

5. SORTING AND CONSERVING DATA

Conservation period

Recommendations:

Not all data needs to be conserved at the end of the project. Certain data should be destroyed if it holds no long-term interest. Other data is conserved for its informational value, potential for reuse, value as evidence, or historic/scientific value.

The University archives department can help you determine what should be done with your data.

The administrative utility period (AUP) is the amount of time that data must be conserved, either because of regulatory constraints, or because it is still necessary and useful for the laboratory(ies) to run smoothly. This period generally starts when a file or project has been closed. The management table from the former Ifsttar archives is a good tool to determine the AUP for data. (<https://intranet.univ-eiffel.fr/luniversite/affaires-juridiques-et-institutionnelles/archivage/documents-utiles-et-fiches-pratiques>)

Exemple de réponse:

- The data will be destroyed after the administrative utility period of 3 years.
- The file listing the names and corresponding identifiers will be deleted at the end of the project.

Archival infrastructure

Recommendations:

Specify how the data will be archived. Will it be conserved in a laboratory storage space? Will it be archived in the infrastructure of the National Computer Centre for Higher Education (CINES) with the help of the Huma-Num major research infrastructure (<https://www.huma-num.fr/services-et-outils/archiver/>)?

6. PUBLISHING DATA AS OPEN RESEARCH DATA

Name of the person(s) responsible for making the data open

Uploading to a data repository

Recommendations:

To make your research data open, it is strongly recommended to upload it to a research repository. There are a large number of such repositories, whether institutional, themed or general. They are listed in the Re3data registry: <https://www.re3data.org/>

Exemple de réponse:

- The project data will be uploaded to the SEANOE marine data portal.
- The project data will be available in the Université Gustave Eiffel institutional space, on the national French repository Recherche Data Gouv (<https://entrepot.recherche.data.gouv.fr/>).

Allocating a permanent ID to the data

Recommandations:

When you upload the data, it is important that a permanent ID be attributed, to make it more accessible and easier to cite in a publication, for example. Make sure that the repository to which the data will be uploaded will allocate an ID (e.g. DOI, Handle). The Recherche Data Gouv repository allocates a DOI to all data uploaded.

Exemple de réponse:

DOI automatically allocated on the Recherche Data Gouv repository.

Date of publication

Recommandations:

Data can be made open at a later date for reasons relating to exploitation and publication, for example.

Exemple de réponse:

The data will be uploaded in the six months following the end of the project.

Licence for reuse

Recommandations:

It is important to give a reuse licence to data that will be open. This will inform people who wish to reuse the data of the conditions for doing so. Université Gustave Eiffel invites people who plan to publish open data to do so under one of the two licences specified by the [Decree n°2017-638 of 27 April 2017 - art. 11](#) that may be used by public bodies: the open licence or the ODbL licence. To find out more, please see the Intranet page "[Maîtriser le cadre juridique](#)" (in French)

Exemple de réponse:

- LO, Etalab open licence
- ODbL

Non-publication

Recommandations:

As a reminder, public data must be open by default. You must therefore explain the reasons for which the data cannot be published. By nature, confidential and sensitive data cannot be published. The same goes for personal data, which can only be published as after it has been made irreversibly anonymous.

Exemple de réponse:

The videos will not be published as open research data as they include data that can be used to identify people. Automatic video anonymisation tools are currently not sufficiently effective and manual anonymisation would take an excessive amount of time.

7. RESOURCES NEEDED FOR DATA MANAGEMENT

Resources

Recommandations:

Data management can involve costs for storage, equipment, and human and financial resources, e.g. to prepare data to be published as open data. In this part, specify the resources planned for data management and publication.

Exemple de réponse:

€300 / TB / year