DMP du projet "Mobiliser et sélectionner la diversité cultivée intra et inter-spécifique pour un changement systémique vers une agriculture sans pesticide"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).

Plan Details

Plan title	DMP du projet "Mobiliser et sélectionner la diversité cultivée intra et inter-spécifique pour un changement systémique vers une agriculture sans pesticide"
Language	fra
Creation date	2021-01-20
Last modification date	2023-04-28
Identifier	PGD_MoBiDiv

Project Details

Project title	Mobiliser et sélectionner la diversité cultivée intra et inter-spécifique pour un changement systémique vers une agriculture sans pesticide
Acronym	MoBiDiv
Abstract	Mobiliser et sélectionner la diversité cultivée intra et inter-spécifique pour un changement systémique vers une agriculture sans pesticide
Funding	• ANR : ANR-20-PCPA-0006

Research outputs :

1. Default research output (Jeu de données)

Contributors

Name	Affiliation	Roles
Jérôme Enjalbert - https://orcid.org/0000-0003-0408-2253		• Personne contact pour les données • Project coordinator
Yannick De Oliveira		• Responsable du plan

Droits d'auteur :

DMP du projet "Mobiliser et sélectionner la diversité cultivée intra et inter-spécifique pour un changement systémique vers une agriculture sans pesticide"

1. Data description and collection or re-use of existing data

1a. How will new data be collected or produced and/or how will existing data be re-used?

The data management and processing (MoBiDiv task 6.2) in MoBiDiv will be based on two pillars: i) FAIR datasets stored in databases, and ii) reproducible research, i.e., versioning of datasets and processing scripts, and their complete description (metadata, documentation). Several datasets will be hosted or generated by MoBiDiv, e.g., analysis of crop diversification on existing datasets (MoBiDiv WP1), simulated datasets by ecophysiological or genetic models (MoBiDiv WP2 & WP5), measures of host-pathogen interactions (MoBiDiv WP2), results from lab or field trials (MoBiDiv WP3), data collected on farm (decentralized evaluation, participatory breeding WP4), recorded interviews (MoBiDiv WP5), databases on key traits for decision support tools (MoBiDiv WP5) and impact-related information (MoBiDiv WP6). All the data produced in the project will be released, after an embargo allowing the publication of the main results. To make data FAIR, MoBiDiv will use different databases and tools already developed or used by the consortium: ShineMas (De Oliveira et al, 2019), ThaliaDB, C
CropDiv, GnpIS, etc. This IT framework will allow (i) the storage of the various data types and (ii) the integrated analysis of these data. The different DB operates with a user database with login, with proper administration of access and editing rights, enabling data upload, quality control, import and download, and a series of management and display tools for data visualization and basic processing–data that will either stay behind login or will be published after quality control.

Data produced in the project will therefore be :

- Observation/Phenotypic data
- Omics (genotyping, transcriptomic ...) data
- Survey data
- Analysis data
- Predictive data

Datasets imported into the project will be accessible to the members in charge of analysis for the project and will respect the term of usage defined in previous project.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

MoBiDiv will produce mainly numeric, textual and audio data. Open and standards formats will be used when possible. Mostly :
CSV or xlsx format for numeric or textual data
mp3 format for audio data
docx format for textual data

Data produced during the project won't exceed 3To.
Spectrophotometric data and genomic data may produce over 1To of data

2. Documentation and data quality

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

- **Metadata**

In a first time datasets documentation will be ensured with the DataCite Metadata standard regarding INRAE Dataverse instance. A working group in MobiDiv will work to evaluate specialized Metadata standard such as Social Science and Humanities standard and Life Sciences standard also proposed in the Dataverse INRAE instance.

Other Metadata standard will be used regarding rules in place in specific data repositories.

- **Data organization**

Each dataset will be organized in a folder with the following structure :

```
|__WPx
    |__data
        |__dataset_ID
            |__ datafile1.csv
            |__ datafile2.csv
            |__ README.txt
```

WP*x* refers to the workpage producing this dataset.

dataset_ID refers to the id in the table attached to the DMP

Each folder will contain a set of files with a REAME file describing the data.

2b. What data quality control measures will be used?

The MoBiDiv consortium will promote the reproducibility of its data analyzes (Goodman et al, 2016). Good practices will be shared among members under the guidance of GQE (Flutre, 2014) along with collaborating teams (MIA-Paris and P3F). Theses will notably involve the usage of free software for statistical analyzes such as R, for simulating with ecophysiological models such as OpenAlea, and for version control such as git hosted on the authentication-based platform SourceSup.

Concerning quality control, data production of different teams on the same information will be alined. Quality check will be performed during data acquisition.

---

3. Storage and backup during the research process

3a. How will data and metadata be stored and backed up during the research?

- **Data stored in NextCloud**

Data and metadata with a low volumetry and concerning the whole consortium or working subgroup will be stored in the INRAE instance of NextCloud, a storage dedicated to MoBiDiv project have been created.

Data will be versionned regarding versioning system provided by NextCloud

The backup of data will be ensured by the IT support of INRAE through NextCloud instance.

- **Data+Script stored in GitLab**

Datasets concerning working groups will be stored in MIA GitLab (https://forgemia.inrae.fr/), backed up and versionned regarding git version control system.

When data are the results of analysis, the scripts computing the analysis will also be versionned using Git.

- **Local Databases and servers**

Large datasets will be stored in locales (UR/UMR) databases or servers. In such cases, data backup is described in the datasets table provided in attachement.

3b. How will data security and protection of sensitive data be taken care during the research

- **NextCloud and Git :**

Each Work Package folder will be shared with the partners of this WP. A dataset folder is accessible to all the members of the WP and can also be shared with members of other WP.
In any cases, data protection will be ensured thanks to the centralized authentication system of INRAE.

- **Local Databases and servers :**

In such cases, security and protection is described in the dataset table provided in attachement.

---

4. Legal and ethical requirements, code of conduct

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

Personal data will be collected in WP 1, 4 and 5 through :

- interviews with stakeholders
- online survey of farmers or stakeholders
- participatory workshop with stakeholders

A mailing list will be set up. We will collect by survey the consent of the stakeholders to be part of this mailing list.

Each workpackage will respect the GDPR.

Before any face-to-face interview or participatory workshop, an informed consent form will be signed by the respondent/participant. Specific consent will be requested for audio/video recordings. The consent form will allow respondents to exercise their right to access, rectify, and object. Consent forms will be kept for five years after the end of the project (i.e., until the end of 2031).

For interviews conducted by telephone, consent will be sought verbally from the respondent and recorded. The audio recordings of the interviews will be kept for the time of their transcription and pseudonymization. Each researcher agrees that the audio recordings will be destroyed at the latest at the end of the project (end 2026). Only anonymized data will be archived.

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Dataset will remain the property of the partnership producing data (INRAE, MNHN, CIRAD, Sorbonne Université, SupAgro).
The details of ownership is accessible in the table provided in attachement and will be agreed in the consortium agreement.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

No ethical issue in MoBiDiv project

---

5. Data sharing and long-term preservation

Most datasets will be accessible after the project (2024) through a public data repository (https://data.inrae.fr/). Data will be open and reusable under citation condition.
Exceptions are made for datasets including personnal data. All datasets under GDPR won't be accessible.

Data will be host on the INRAE instance of Dataverse, a public data repository. All data will be preserved as long as INRAE maintain this service.

Using data do not need a specific software.

A DOI will be assigned to each dataset (generated by the Dataverse platform itself) and each one will be accessible with the Dataverse API.

6. Data management responsibilities and resources

Each data producer will be responsible of its own data management.
Yannick De Oliveira will be responsible of the DMP and the data stewart during MoBiDiv project.

4 non-permanent person.month and 2 permanent (Yannick De Oliveira) person.month will be dedicated to the data management during the project.