

---

# DMP du projet "Mettre le plan de gestion des données automatisé dans les mains des biologistes"

*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

<b>Plan title</b>	DMP du projet "Mettre le plan de gestion des données automatisé dans les mains des biologistes"	
<b>Version</b>	Final version	
<b>Fields of science and technology (from OECD classification)</b>		
<b>Language</b>	fra	
<b>Creation date</b>	2020-06-22	
<b>Last modification date</b>	2022-09-15	
<b>Identifier</b>	maDMP4LS	
<b>License</b>	<b>Name</b>	Creative Commons Attribution 4.0 International
	<b>URL</b>	<a href="http://spdx.org/licenses/CC-BY-4.0.json">http://spdx.org/licenses/CC-BY-4.0.json</a>
<b>Management plans related to the project</b>	<ul style="list-style-type: none"><li>Initial version of the DMP with a non structured model : <a href="https://dmp.opidor.fr/plans/6580">https://dmp.opidor.fr/plans/6580</a></li></ul>	

## Project Details

**Project title** Mettre le plan de gestion des données automatisé dans les mains des biologistes

**Acronym** maDMP4LS

**Abstract** L'Institut Français de Bioinformatique (IFB), avec ses deux infrastructures informatiques et ses 30 plates-formes, est une structure essentielle pour les Sciences de la Vie et ses applications (santé, agriculture, environnement), fournissant un environnement de production, d'analyse et de gestion de données pour les communautés d'utilisateurs. Bien que ces communautés puissent être considérées comme bien dotées d'outils de structuration et de gestion des données, des problèmes subsistent pour mettre pleinement en œuvre les bonnes pratiques et gérer les données, de leur production à leur préservation, en assurant leur accessibilité, leur reproductibilité et leur réutilisation. Le Plan de Gestion des Données (PGD ou DMP pour Data Management Plan) est considéré comme un élément-clé qui facilitera la mise en œuvre des principes FAIR (Findable, Accessible, Interoperable, Reusable). Le DMP contient les éléments qui guident le processus de gestion des données tout au long de leur cycle de vie et fournissent également des traces de la provenance des données (inclus tous les résultats de recherche dérivés des données brutes). L'Inist-CNRS propose l'outil Web DMP-OPIDoR qui facilite la rédaction des DMP et contribue également à l'harmonisation des bonnes pratiques en fournissant conseils, exemples et modèles spécifiques de certaines institutions. L'évolution du DMP OPIDoR vers un DMP automatisable (« machine-actionable DMP ») est actuellement à l'étude sur la base des retours d'expérience des utilisateurs, des cas collectés et des travaux de RDA sur le DMP actif. L'outil DMP OPIDoR « machine actionable » sera conforme au modèle commun RDA DMP mais inclura également une extension afin de servir les différents acteurs et besoins disciplinaires validés par un large groupe d'utilisateurs.

**Funding**

- Agence nationale de la recherche (ANR) : ANR-19-DATA-0017-01

**Partners**

- Institut de l'information scientifique et technique
- Institut français de bioinformatique  
<https://ror.org/045f7pv37>

**Research outputs :**

1. machine actionable DMP OPIDoR (Software)
2. machine actionable DMP compatible user and project management system for Bioinformatics (Software)
3. Formation à l'utilisation de maDMP4LS (Text)

**Contributors**

Name	Affiliation	Roles
Bourhy Konogan - <a href="https://orcid.org/0000-0001-6234-1263">https://orcid.org/0000-0001-6234-1263</a>	Institut Français de Bioinformatique - 045f7pv37	
Faure Benjamin - <a href="https://orcid.org/0000-0003-4521-9252">https://orcid.org/0000-0003-4521-9252</a>	INIST - 02mn0vt57	
Jacquemot-Perbal Marie-Christine - <a href="https://orcid.org/0000-0002-6316-1472">https://orcid.org/0000-0002-6316-1472</a>	INIST - <a href="https://ror.org/02mn0vt57">https://ror.org/02mn0vt57</a>	<ul style="list-style-type: none"> <li>• DMP manager</li> <li>• Personne contact pour les données (maDMP OPIDoR)</li> </ul>
Olivier Collin - <a href="https://orcid.org/0000-0002-8959-8402">https://orcid.org/0000-0002-8959-8402</a>	IRISA - 00myn0z94	<ul style="list-style-type: none"> <li>• Personne contact pour les données (maTraining, maMY)</li> </ul>
VAN HELDEN Jacques - <a href="https://orcid.org/0000-0002-8799-8584">https://orcid.org/0000-0002-8799-8584</a>	Institut français de bioinformatique - 201321719F	<ul style="list-style-type: none"> <li>• Project coordinator</li> </ul>

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

# DMP du projet "Mettre le plan de gestion des données automatisé dans les mains des biologistes"

---

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

### machine actionable DMP OPIDoR

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

Developments will be made on the existing DMP OPIDoR code that is available and accessible in a Github repository: <https://github.com/OPIDoR/DMPOPIDoR> under an MIT license. This code is itself based on DMP roadmap code (<https://github.com/DMPRoadmap>) and new features will, if relevant, be integrated in DMP OPIDoR code.

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

The application is developed using Ruby on Rails framework that implements a Model-View-Controller approach. REST API endpoints will be developed.

### machine actionable DMP compatible user and project management system for Bioinformatics

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

[my.genouest.org](https://my.genouest.org) is an account and project management developed by GenOuest. The source code is available at <https://github.com/genouest/genouestaccountmanager>.

A new branch code will be created for the maDMP-OPIDoR compatible version.

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

The new version will be available in the same repository : <https://github.com/genouest/genouestaccountmanager>

### Formation à l'utilisation de maDMP4LS

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

maTraining consists of training materials that will be used for "train the trainer" sessions. The training materials will

consists of documentation and tutorials.

---

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

Training materials will be made available online on open-access repositories like zenodo. The training material will consist of presentations (pdf), text documents (pdf) and possibly videos.  
The volume will be rather low: a few gigabytes.

## 2. Documentation and data quality

### **machine actionable DMP OPIDoR**

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

Documentation is organized as Markdown files and is available in a Github depot :  
<https://github.com/OPIDoR/DMPOPIDoR>. The documentation includes installation instructions.

Metadata compliant with DataCite metadata schema will be produced so as to facilitate the discovery of the code and its citation.

---

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

A Jenkins server was deployed and continuous integration will be applied to build the software and run the tests.

### **machine actionable DMP compatible user and project management system for Bioinformatics**

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

Documentation is organized as Markdown files and is available in a Github depot :  
<https://github.com/genouest/genouestaccountmanager>  
The documentation includes installation instructions.

---

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

The code is checked in Continuous Integration manner with Travis-CI.

## Formation à l'utilisation de maDMP4LS

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

Question sans réponse.

---

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

The content of the training will be evaluated internally by giving the training to IFB members, thus allowing to have their feedback before giving the training to end-users.

## 3. Storage and backup during the research process

### machine actionable DMP OPIDoR

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

The generated code is stored on local servers for development purposes. The local machines are backed up on Inist infrastructure. The source code is also stored on Gitbucket with restricted access and an open Github depot.

---

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

No sensitive data are handled.  
REST APIs will integrate security patterns and be compliant with GDPR.

### machine actionable DMP compatible user and project management system for Bioinformatics

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

The code is stored on local workstations during development phases. The workstations are backed up on Inria/Irisa infrastructure. The source code is also stored on the Github repository.

---

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

The My environment is used for user management on computing facilities. It is used as a front server for a more convenient management of LDAP information and adds many features for project management as well as features for end users such as password recovery and authentication keys management.

The administrative access is restricted to site administrators. The connexion of an administrator needs a two factor authentication in order to block unauthorized access.

The LDAP directory is backed up on Irisa's server and the My database is backed up on GenOuest infrastructure.

## **Formation à l'utilisation de maDMP4LS**

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

The training material will be hosted on public repositories.

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

Question sans réponse.

## **4. Legal and ethical requirements, code of conduct**

### **machine actionable DMP OPIDoR**

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

These data do not contain any personal data.

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

The initial DMP roadmap code was published under an MIT license and so will be the source code of DMP OPIDoR. This license is included in the list established through the French Law for a Digital Republic.

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

Question sans réponse.

## **machine actionable DMP compatible user and project management system for Bioinformatics**

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

No personal data processing.

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

The original My source code is under a GNU Affero General Public License v3.0. So will be the maMy code branch.

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

No ethical issues.

## **Formation à l'utilisation de maDMP4LS**

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

No personal data are processed.

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

Question sans réponse.

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

Question sans réponse.

## 5. Data sharing and long-term preservation

### machine actionable DMP OPIDoR

#### 5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

The source code will be deposited on Github. No embargo will be applied.

#### 5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

A version of the source code will be deposited in HAL.

#### 5c. What methods or software tools are needed to access and use data?

The code is openly accessible on Github. Anyone is free to clone or download the code and reuse it, modify it.

#### 5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

A DOI will be attributed to a version of the source code.

### machine actionable DMP compatible user and project management system for Bioinformatics

#### 5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

The source code will be deposited on Github. No embargo will be applied.

#### 5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

Long term preservation will be achieved thanks to Software Heritage that harvest Github repositories for long term archival. Archive are associated with unique identifiers (SWHID).

**5c. What methods or software tools are needed to access and use data?**

The code is openly accessible on Github. Anyone is free to clone or download the code and reuse it, modify it.

**5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**

Thanks to Software Heritage the applicate code repository has a unique identifier

**Formation à l'utilisation de maDMP4LS**

**5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?**

The training support material is openly accessible.

**5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?**

Training material will be accessible on the github repository.

**5c. What methods or software tools are needed to access and use data?**

Any web browser can be used.

**5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**

Since github is harvested by software heritage, a unique identifier (SWHID) will be associated to the training resource.

**6. Data management responsibilities and resources**

**machine actionable DMP OPIDoR**

**6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**

DMP-OPIDoR team : Florian Mazur (Software project leader), Benjamin Faure (Software developer), Jean-Michel Parret (Head of service) , Françoise Cosserat (User test coordinator), Anne Busin (User test), Laurent Rassinoux (User test), M-Christine Jacquemot (Product owner)

---

**6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**

Question sans réponse.

---

## **machine actionable DMP compatible user and project management system for Bioinformatics**

**6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**

GenOuest / IFB maDMP team :

- Konogan Bourhy : developer
- Olivier Sallou : original developer of my.genouest.org
- Mateo Boudet : former developer of project module of my.genouest.org
- Olivier Collin : team leader
- Jacques van Helden : director of IFB

---

**6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**

No specific resources since all the developments are immediately available on the repositories.

---

## **Formation à l'utilisation de maDMP4LS**

**6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**

There will be a joint responsibility between IFB and Inist for the training since both Inist and IFB propose training and tutorial sessions for their tools.

---

**6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**

Training material will be elaborated by all IFB members involved in the development of training material. Thanks to the interaction with ELIXIR, training will be made available in the TESS resource.

