

maDMP4LS

Machine Actionable DMP for Life Sciences

IFB - Inist

Konogan Bourhy, Olivier Collin,
Marie-Christine Jacquemot, Florian Mazur, Benjamin Faure,
Anne Busin, Françoise Cosserat, Laurent Rassinoux,
Jean-Michel Parret

Who are we?

- Development Engineer
- Joined the Genouest team in September 2020

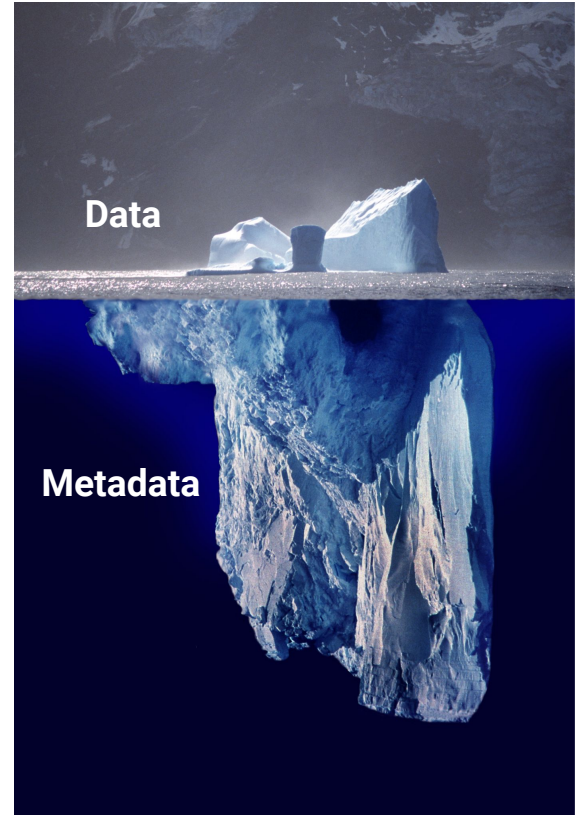


- Located at the IRISA lab in Rennes, on the Beaulieu campus
- Bioinformatic facility, giving access to informatic tools to researchers.
- Several solutions:
 - Cluster
 - Web portal to access tools from Galaxy
 - Virtual environments using Genostack
- Cesgo tools for collaboration

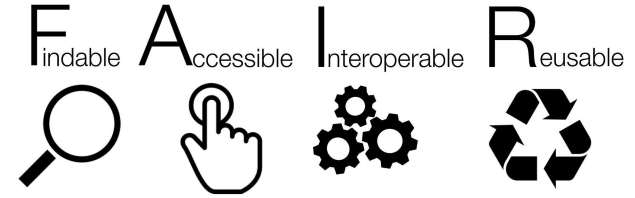
<https://www.genouest.org/>
<https://www.cesgo.org/fr/>

The push for Open science

- Never been more data generated
- Huge quantities of data stored
 - even more metadata
- How could researchers easily share their work or build upon the work of others?



FAIR data



- General concept, independent of the type of (meta)data
- Findable: Make sure that the data is easily findable for **both humans and computers**
- Accessible: Make sure the data is **accessible** and **safe**
- Interoperable: Uses international standards and vocabularies to **integrate** with other data and processes
- Reusable: Data should be **well-described** (metadata) so it can be replicated and/or combined in different settings

ANR call for projects in 2019

Open Science : research practices and open research data

Goal: Tackle the emerging need to accelerate the adoption of practices for accessibility, reuse and openness of research data.

What do we want to tackle?

- Researcher A wants to adapt researcher B experiences to researcher C datasets
 - Lack of metadata
 - Not structured
- Researcher A then wants to work with a bioinformatics facility
 - Has to explain his needs
 - How long will the project last?
 - Storage cost and ecological impact

“Need bioinformatic tools to carry out analysis”
“For work” “I need the same environment as colleague X”
- Help researchers spend more time doing actual research
 - Less administrative work

What if all these issues could be solved using a single solution...

- Consortium between IFB and Inist proposed “machine actionable DMP for Life Sciences”
- Data Management Plans or DMP are asked for by Funding agencies
- Describe the data generated by the future project

Objective: Transforming the DMP file into a machine actionable data structure

The project started in March 2020 for 18 months (ANR-19-DATA-0017-01)

Who is involved

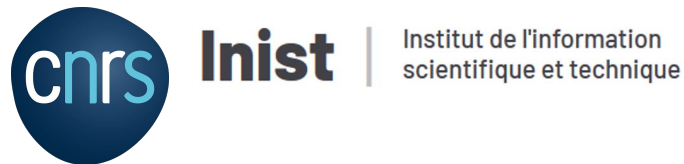
IFB : Institut Français de Bioinformatique / French Bioinformatics Institute

- National Network of Computing resources (NNCR)
- 21 bioinformatics facilities in France

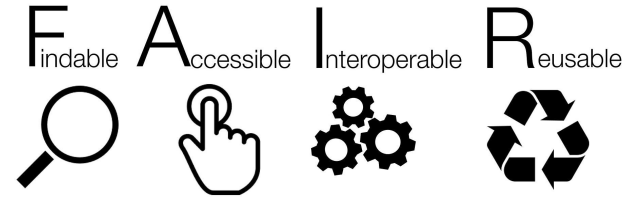


Inist : Institut de l'Information Scientifique et Technique

- Provider of OPIDoR tools (Optimiser le Partage et l'Interopérabilité des Données de la Recherche)
 - Cat OPIDoR : identifying information facilities in France
 - PID OPIDoR : Digital Object Identifier allocation service
 - DMP-OPIDoR : DMP online editor



From the DMP to the machine actionable DMP



- Produce a structured and standardized DMP content:
 - To keep a common data model with different DMP tools
 - To allow automatic systems to act throughout the data life cycle
- Use of internal/external registries and information systems :
 - to pre-populate DMP by getting informations from the financer (ex: ANR)
 - to guide users through the selection of standards, or repositories, tools, etc. (FAIR principles)
- Accessible editor for the researcher to create and edit his DMP during his project

What is DMP OPIDoR?

Online tool allowing the redaction of DMPs

Based on DMP Roadmap 

Adapted to meet the French community needs:

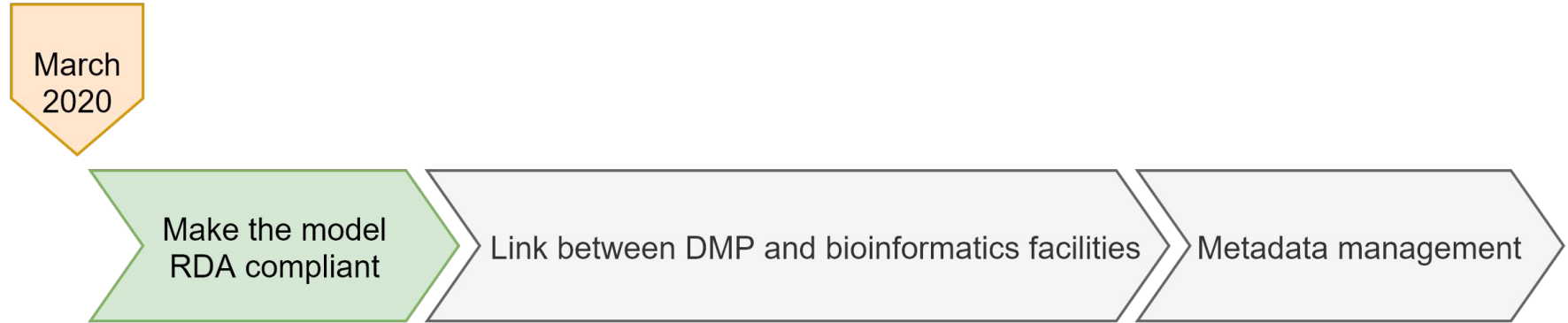
- Easy use of templates
- Edition features
- Compliance with GDPR
- Growing community and ecosystem

DMP : 5735
Templates : 37
Users : 6400



The screenshot displays the 'Rédiger' (Edit) tab of the DMP OPIDoR tool. The interface is in French and includes a navigation bar at the top with links: Informations générales, Produits de recherche, Modèle choisi, Rédiger, Partager, Demande d'assistance conseil, and Télécharger. The main content area is titled 'Description des données et collecte des données ou réutilisation de données existantes'. Below this title, there are two expandable sections: 'tout développer | tout réduire'. The first section, 'Description générale du produit de recherche', contains three questions: 'Description générale du produit de recherche', 'Est-ce que des données existantes seront réutilisées ?', and 'Comment seront produites/collectées les nouvelles données ?'. The second section, 'Documentation et qualité des données', contains two questions: 'Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?' and 'Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?'. A third section, 'Exigences légales et éthiques, code de conduite', contains two questions: 'Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?' and 'Quelles sont les contraintes juridiques (sensibilité des données autres qu'à caractère personnel, confidentialité, ...) à prendre en compte pour le partage et le stockage des données ?'. Each question is followed by a right-pointing arrow indicating it can be expanded.

Project structuration



DMP-OPIDoR data model evolution

Methodology

Take into account:

- RDA DMP Common Standards work
- DMP templates that are published in DMP OPIDoR
- User stories requiring information exchange



Currently, exchange with, and collection of feedbacks coming from different types of services (Funding agencies, computing centres, data providers, researchers, etc.)

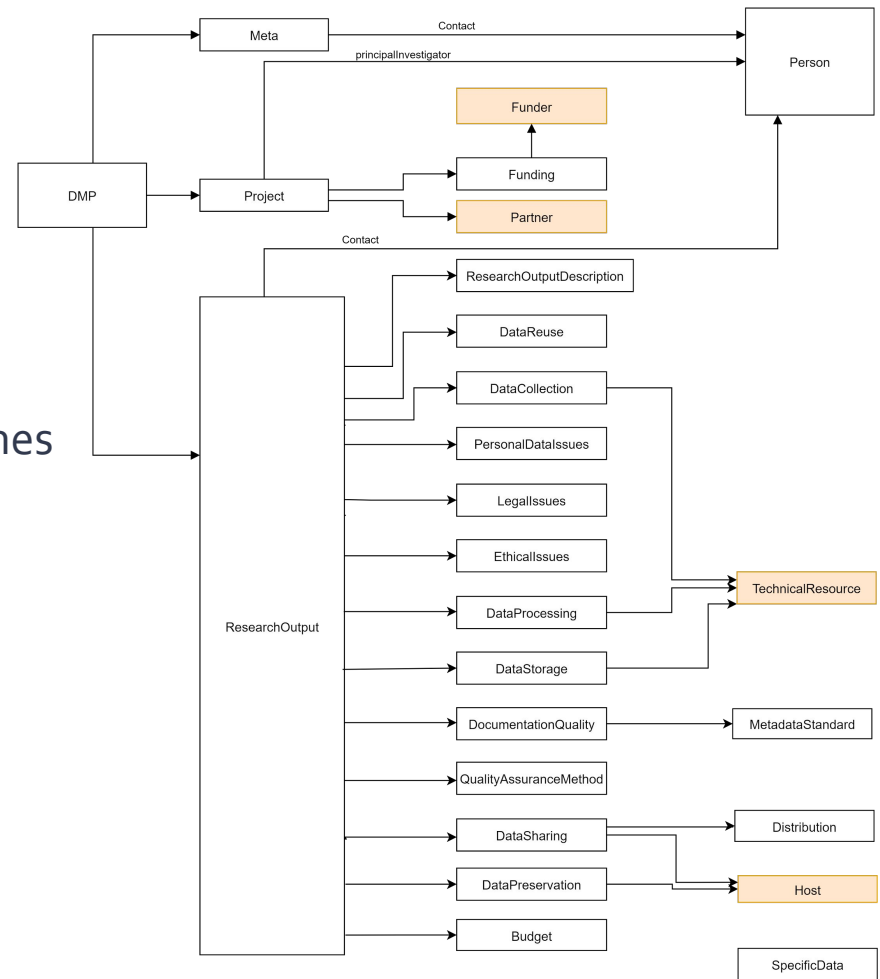
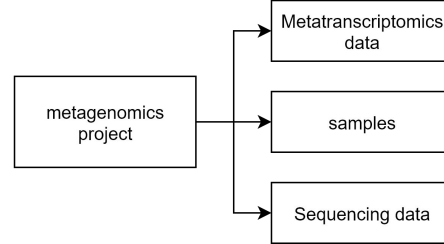
Output

semi-flexible and extensible data model:
adaptation to disciplinary or service specificities

Model overview

Top-level entries :

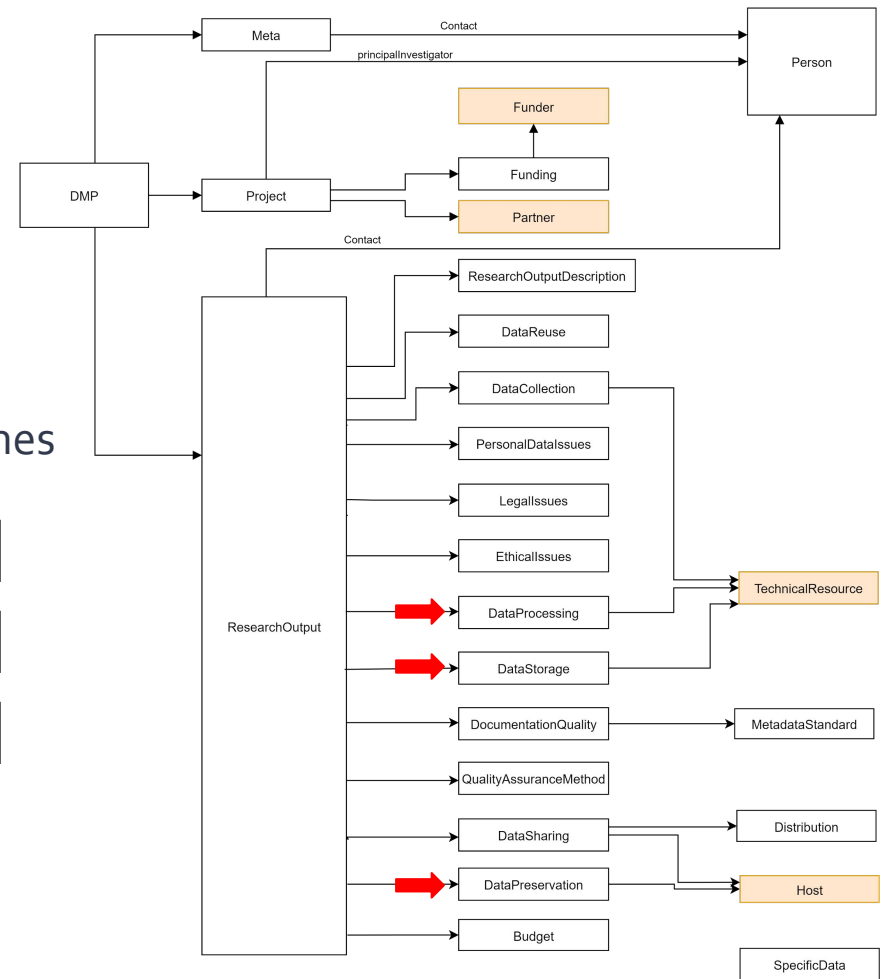
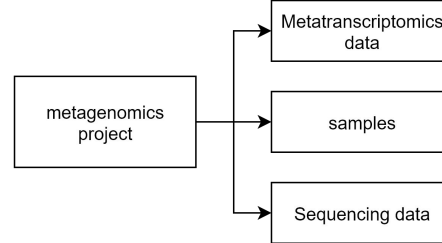
- Meta : metadata on the project
- Project : info about the project
- researchOutput : one or several outcomes of the project



Model overview

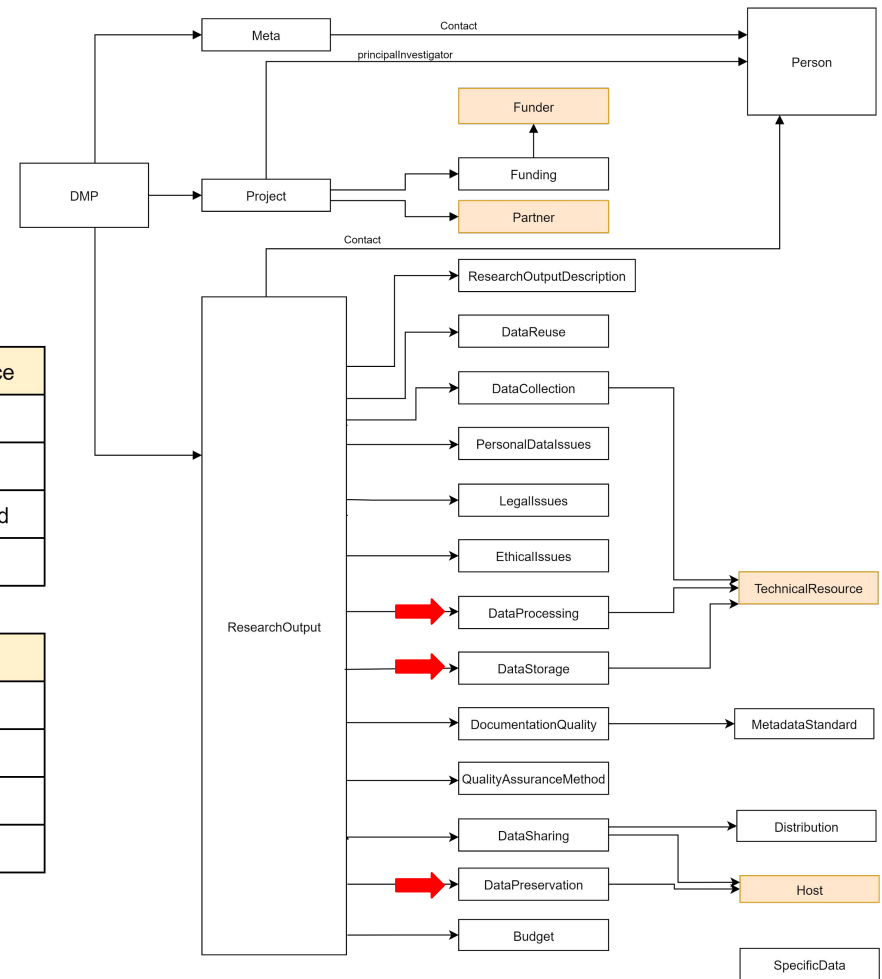
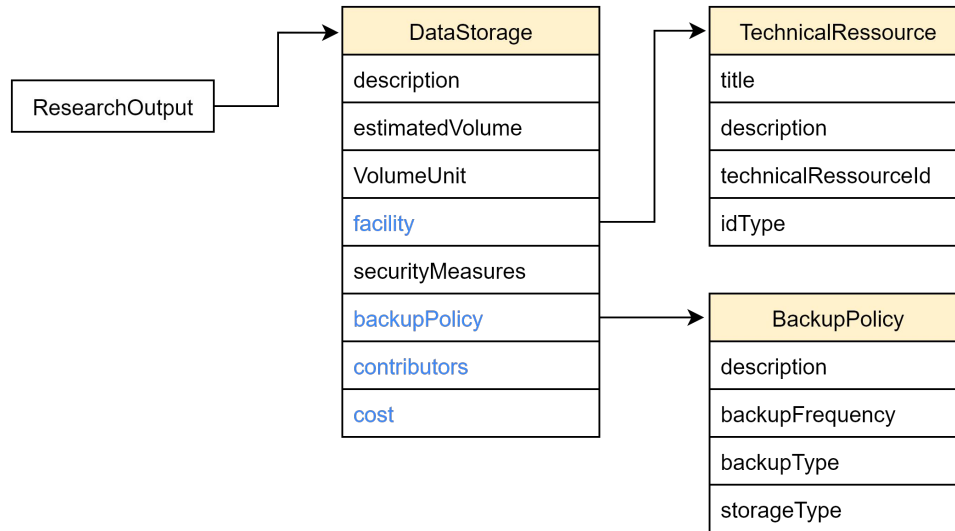
Top-level entries :

- Meta : metadata on the project
- Project : info about the project
- researchOutput : one or several outcomes of the project

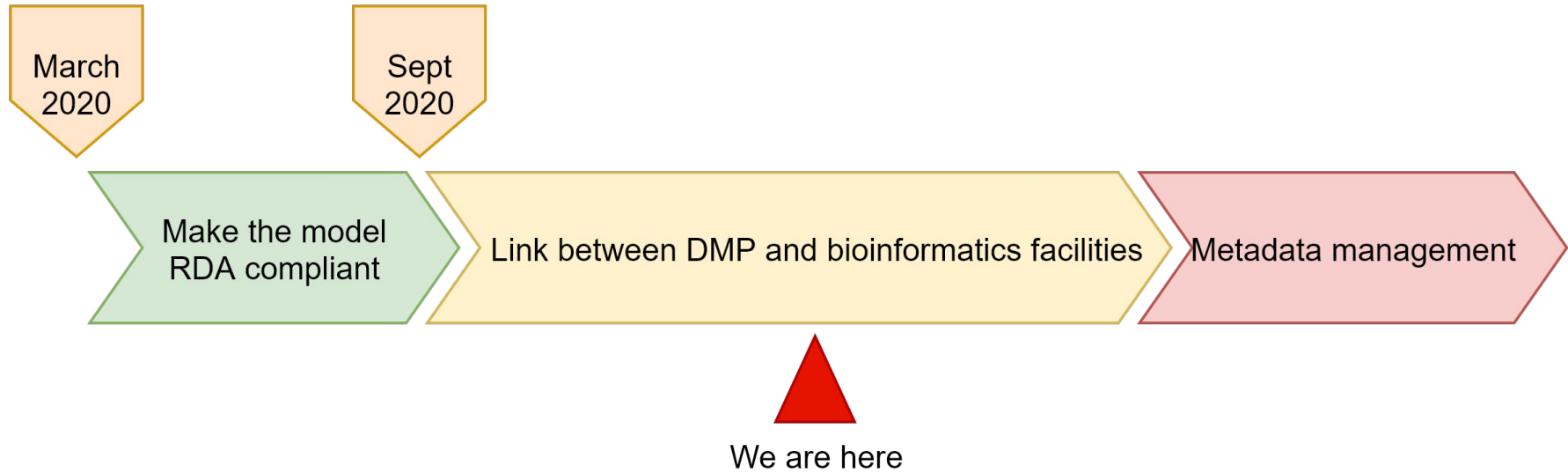


Model overview

Deeper levels of structuration



Project structuration



How should bioinformatics facilities integrate maDMPs?

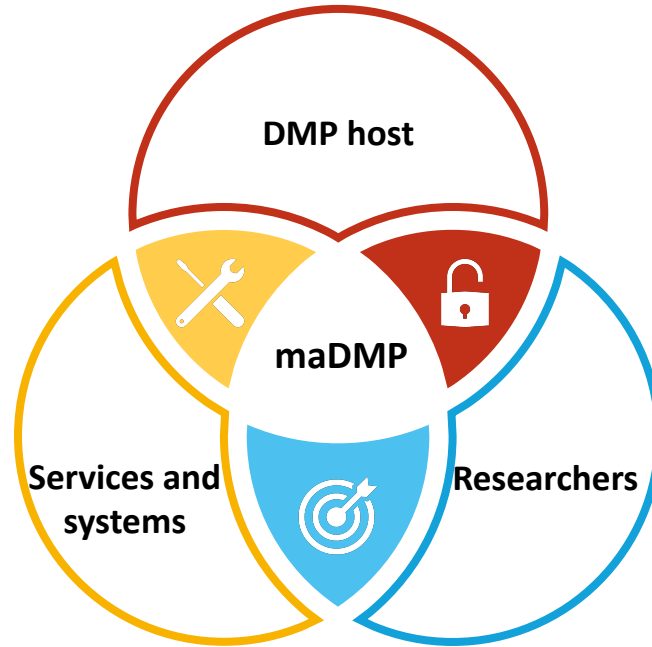
- Pilot project: Genouest to act as stakeholder
- Adapt our tools to handle DMPs and setup the ressources asked for
 - Seamless transition from the “by hand” way to the machine controlled one
 - Only a validation from the Administrators
- Ensure that all necessary data is in the DMP
 - Data included in the standard model
 - Use templates to ask for additional informations
- Additional behaviors may be added to better share informations about the ongoing project
 - Update the DMP
 - Notifications

The screenshot displays the DMP-ODiR web application. The top navigation bar includes links for 'Tableau de bord', 'Créer des plans', 'DMPs publics', 'Modèles de DMP', 'Aide', and 'Plus'. The user is logged in as 'DMP Administrator'. The main content area is titled 'Description' and contains a text editor with the following text: 'Des demandes d'espaces de calcul, analyses sont soumises à la plateforme GenOuest'. Below this is a section titled 'Identification de la ressource' with fields for 'Titre' (GenOuest), 'Description' (a detailed paragraph about GenOuest's infrastructure and services), 'Identifiant de la ressource technique' (https://www.genouest.org/), and 'Adresse contact du service' (https://www.genouest.org/contact). On the right side, there is a 'Forme' dropdown menu set to 'DataStorageGenOuest' and a 'Commentaires' section. At the bottom, there are two buttons: 'Notifier GenOuest' and 'Enregistrer'.

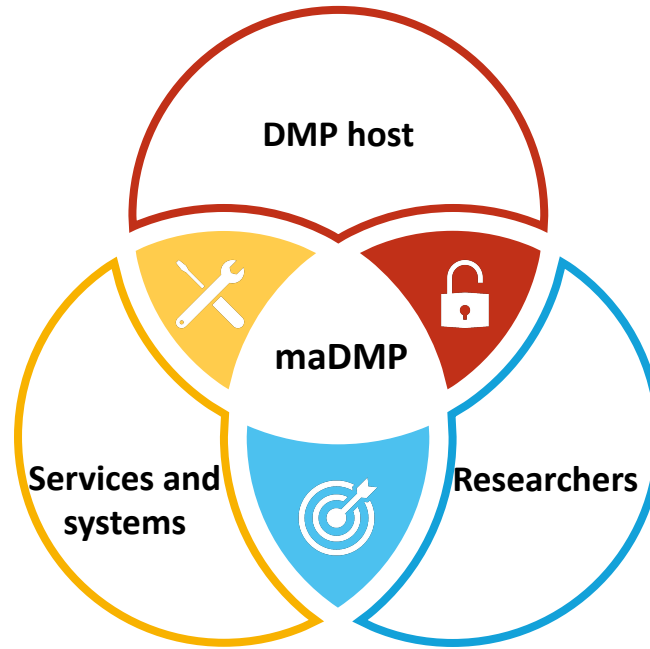
maDMP for researchers

- From an administrative chore to an asset
 - Fully integrated in the research process
- Associated to a Digital Object Identifier
 - Traceability of both data and processes involved in the project
- Give control of the data to the researcher
 - Shareable with trusted stakeholders
 - Ability to define what is accessible in the DMP

The benefits of a machine actionable DMP



The benefits of a machine actionable DMP

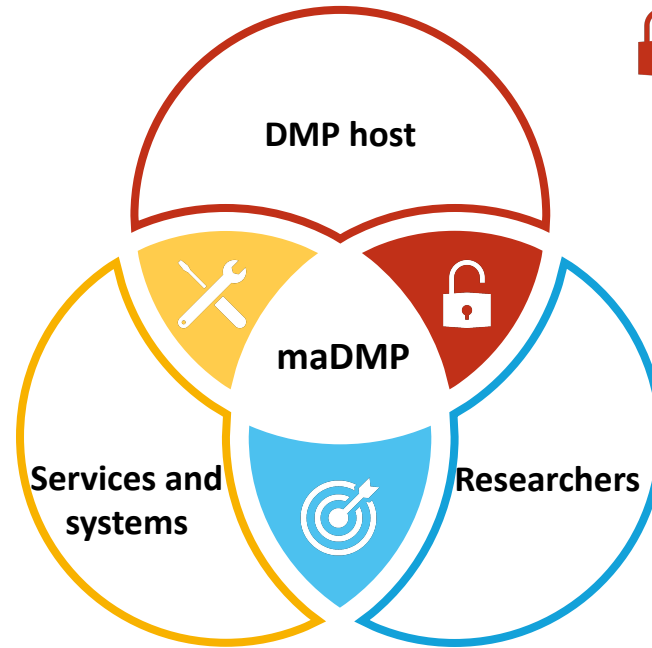


Communication



- Understanding the specific needs of the researcher
- Less work for the researcher (avoiding double capture)
- Better visibility for facilities

The benefits of a machine actionable DMP



Communication



- Understanding the specific needs of the researcher
- Less work for the researcher (avoiding double capture)
- Better visibility for facilities



FAIR access

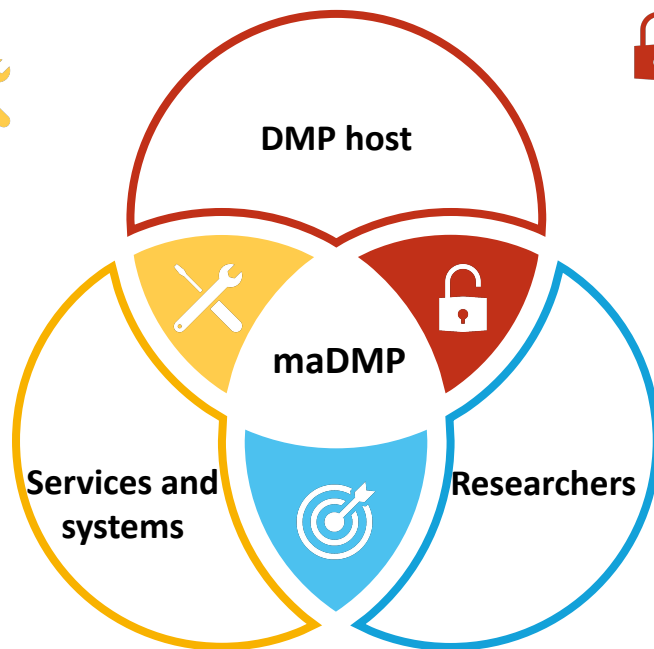
- Fostering safe FAIR principles application
- Share as much as possible
- Protect the data as much as needed
- Accuracy of the DMP

The benefits of a machine actionable DMP

Data Management



- Better management of resources
- DMP up to date
- Ensuring FAIRness of data



Communication



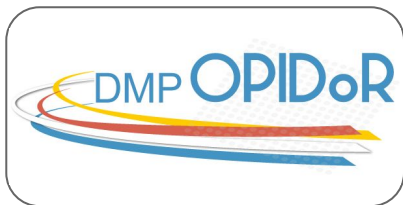
- Understanding the specific needs of the researcher
- Less work for the researcher (avoiding double capture)
- Better visibility for facilities






FAIR access


- Fostering safe FAIR principles application
- Share as much as possible
- Protect the data as much as needed


The future data life cycle



1. Start writing the DMP

 [Tableau de bord](#) **[Créer des plans](#)** [DMPs publics](#) [Modèles de DMP](#) [Aide](#) [Plus](#)  [Français](#)  [DMP Administrator](#)

Science Europe  Admin

 Vous utilisez une application de test. Les modèles disponibles sont des modèles de test. Vous pouvez accéder à la version de production à l'adresse suivante : <https://dmp.opidor.fr/>

Titre du projet


☐ projet de test, d'entraînement ou à des fins de formation

Choisissez un modèle

Vous pouvez choisir soit un modèle fourni par votre organisme soit par un autre organisme, ou un modèle financeur. Le modèle par défaut est **Science Europe** : **modèle structuré standard**.
Retrouvez la liste des modèles disponibles

Science Europe (Votre organisme) **Autre organisme** [Financeur](#)

Plusieurs modèles sont disponibles, lequel souhaitez-vous utiliser ?



Veillez sélectionner un modèle dans la liste.

[Créer un plan](#) [Suivant](#) [Utiliser le modèle par défaut](#)

The future data life cycle

2. Get the project informations





1. Start writing the DMP

DMP OPIDoR My Dashboard Create plans Public DMPs DMP Templates Help More ▾ English(

Indicate the funding of the project


: ANR-07-BDIV-0003



Create


Project start date

jj/mm/aaaa



Project end date

jj/mm/aaaa



List the partners associated with the project

Select a value you want to add to your plan or type a new one.

Indicate the experimental plan reference

Reload

Save

New data have been added to your plan, please click on the "Reload" button.

The future data life cycle

2. Get the project informations



1. Start writing the DMP

DMP OPIDoR My Dashboard Create plans Public DMPs DMP Templates Help More

Vous utilisez une application de test. Les modèles disponibles sont des modèles de test.
Vous pouvez accéder à la version de production à l'adresse suivante : <https://dmp.opido.fr/>

Phytoclim

General informations Research outputs Selected Template Write Plan Share Download

Project Details

☐ mock project for testing, practice, or educational purposes

Project title
Modelling climate impact on an emerging disease, the Phytophthora alni-induced alder dieback

Project acronym
PHYTOCLIM

Project abstract
From the 1990s, a new disease appeared on alder trees along rivers. It is caused by a pathogenic fungus, Phytophthora alni, which attacks tree roots and causes them to die off. This pathogen is a thermophilic species. Moreover, the emergence of alder dieback is concomitant with the particularly hot years of the last two decades. Environmental factors, such as mean site temperature and soil characteristics, play an important role in the occurrence of the disease. The objective of the present work was to model and forecast the effect of environment on the spread of alder Phytophthora infection...

Plan Guidance Configuration
To help you write your plan, DMP OPIDoR can show you guidance from a variety of organisations.
Select up to 6 organisations to see their guidance.
☐ Digital Curation Centre - DCC
Find guidance from additional organisations below
[See the full list](#)
Save

Fill informations if ANR project **Save**

DMP OPIDoR My Dashboard Create plans Public DMPs DMP Templates Help More

Indicate the funding of the project

Funder: funding identifier	Actions
French National Research Agency : ANR-07-BDIV-0003	✎ ✕

Create

Project start date
jj/mm/aaaa

Project end date
jj/mm/aaaa

List the partners associated with the project
Select a value you want to add to your plan or type a new one.

Biodiversité, Gènes et Communautés (200317684N)	👁 ✕
Walloon Agricultural Research Centre ()	👁 ✕
Plant Protection Institute ()	👁 ✕

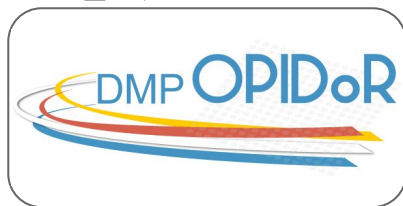
Indicate the experimental plan reference

Resource title

Fill informations if ANR project **Save**

The future data life cycle

2. Get the project informations



3. Get the project going



1. Start writing the DMP

Request a new project creation

DmpID (Optional)

Allows us to link your project to your DMP hosted by DMP OPIDoR

Name (required)

Avoid generic name, team name, technology name or your name. Please, choose a project name that matches your cluster research project. If you treat several projects, it is quite possible for you to request more project spaces.

Size (GB)

Optional, for information only

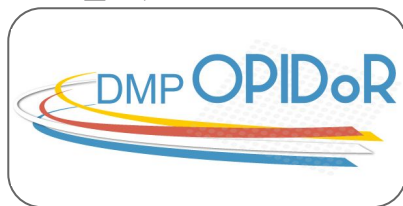
Financing

Optional, for information only. If several organisms are financing your project, please separate the names with a comma (Example: CNRS,INRAE,INSERM)

Description

The future data life cycle

2. Get the project informations



3. Get the project going



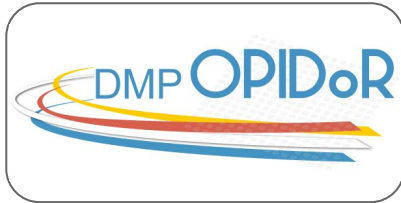
1. Start writing the DMP

Projects						
Project	Owner	Group	Path	Size (Go)	Creation	Expiration
phytoclim	kbourhy	cnrs	/opt/project /phytoclim	600	2021-05-10	2022-05-05
1 of 1 << < 1 > >> 10 ▾						

- Workspace specificities automatically established
- Collaborators with a Genouest account added

The future data life cycle

2. Get the project informations



3. Get the project going



1. Start writing the DMP

Projects						
Project	Owner	Group	Path	Size (Go)	Creation	Expiration
phytoclim	kbourhy	cnrs	/opt/project/phytoclim	600	2021-05-10	2022-05-05

1 of 1 << < 1 > >> 10 ▾

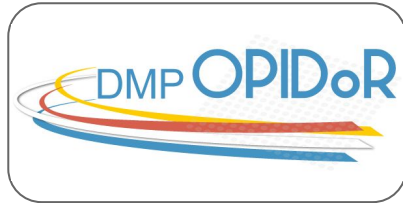


Projects						
Project	Owner	Group	Path	Size (Go)	Creation	Expiration
phytoclim	kbourhy	cnrs	/opt/project/phytoclim	2000	2021-05-10	2022-05-05

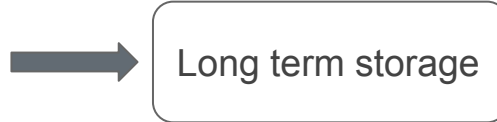
1 of 1 << < 1 > >> 10 ▾

The future data life cycle

2. Get the project informations

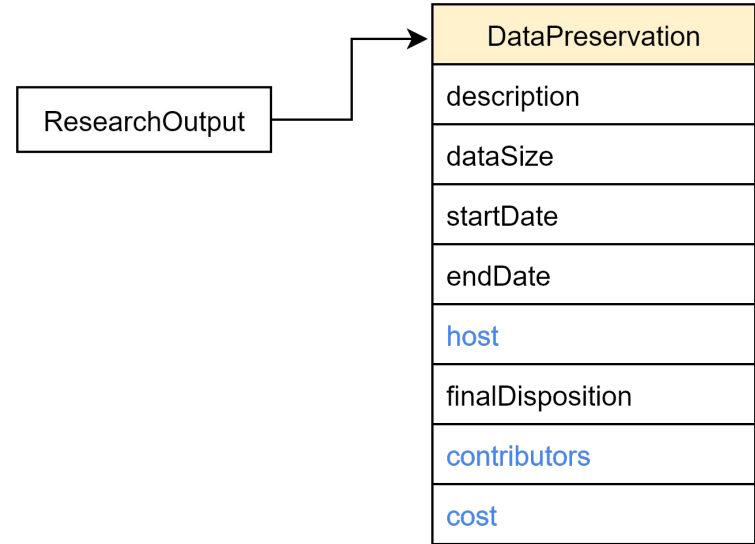


3. Get the project going



1. Start writing the DMP

4. Store the data for long term



Acknowledgments

Genouest team:

Olivier Collin

Opidor team:

Marie-Christine Jacquemot

Florian Mazur

Benjamin Faure

Françoise Cosserat

Laurent Rassinoux

Jean-Michel Parret

Anne Busin

<https://www.genouest.org/>

<https://dmp.opidor.fr/>