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?

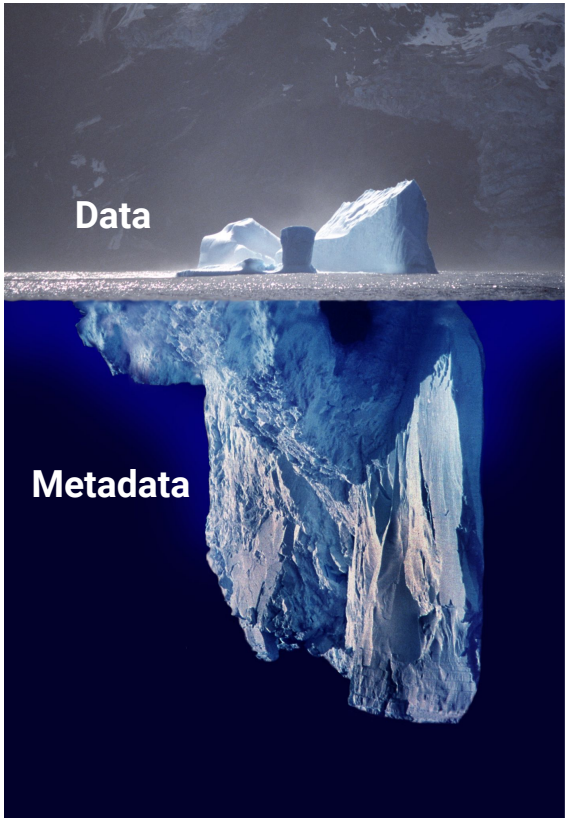
- Development Engineer
- Joined the Genouest team in September 2020

- 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

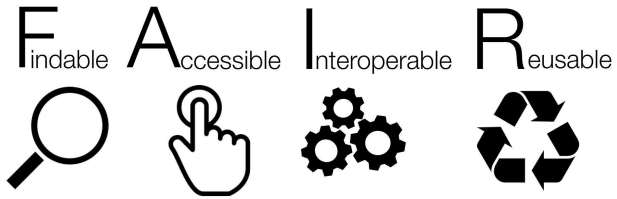
https://www.genouest.org/
https://www.cesgo.org/fr/

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

https://www.go-fair.org/fair-principles/

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…

maDMP4LS

- 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

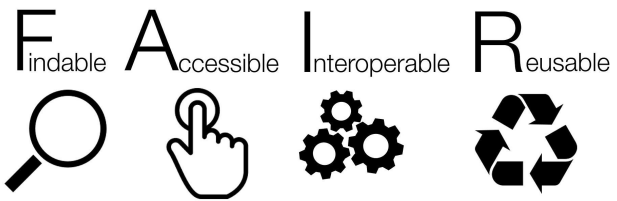
IFB : Institut Français de Bioinformatique / French Bioinformatics Institute

- 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

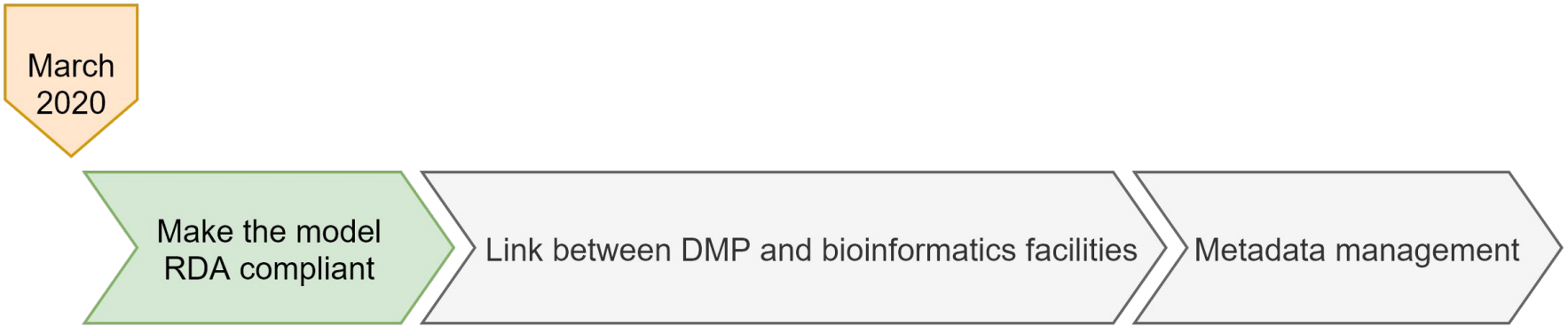
Based on DMP Roadmap roadmap

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

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.)

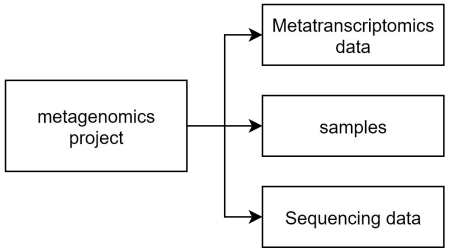
Output

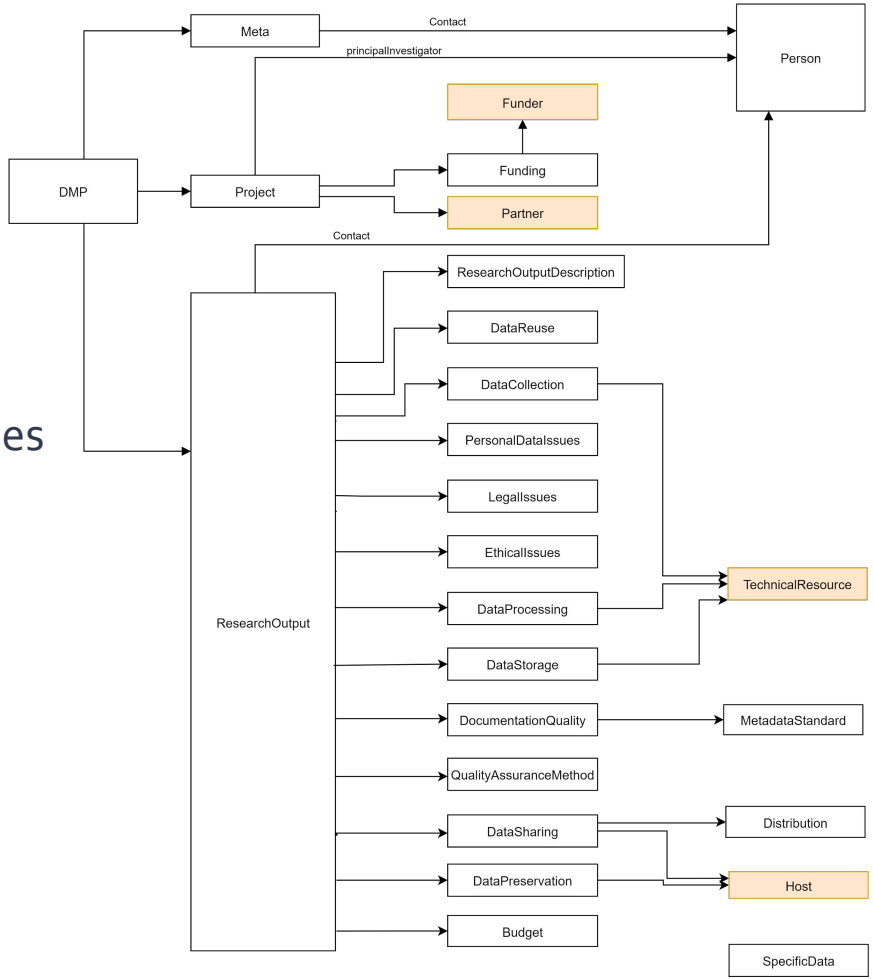
semi-flexible and extensible data model:
adaptation to disciplinary or service specificities

Model overview

Top-level entries :

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

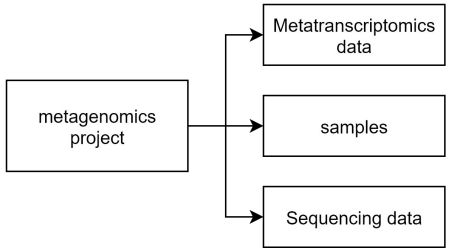


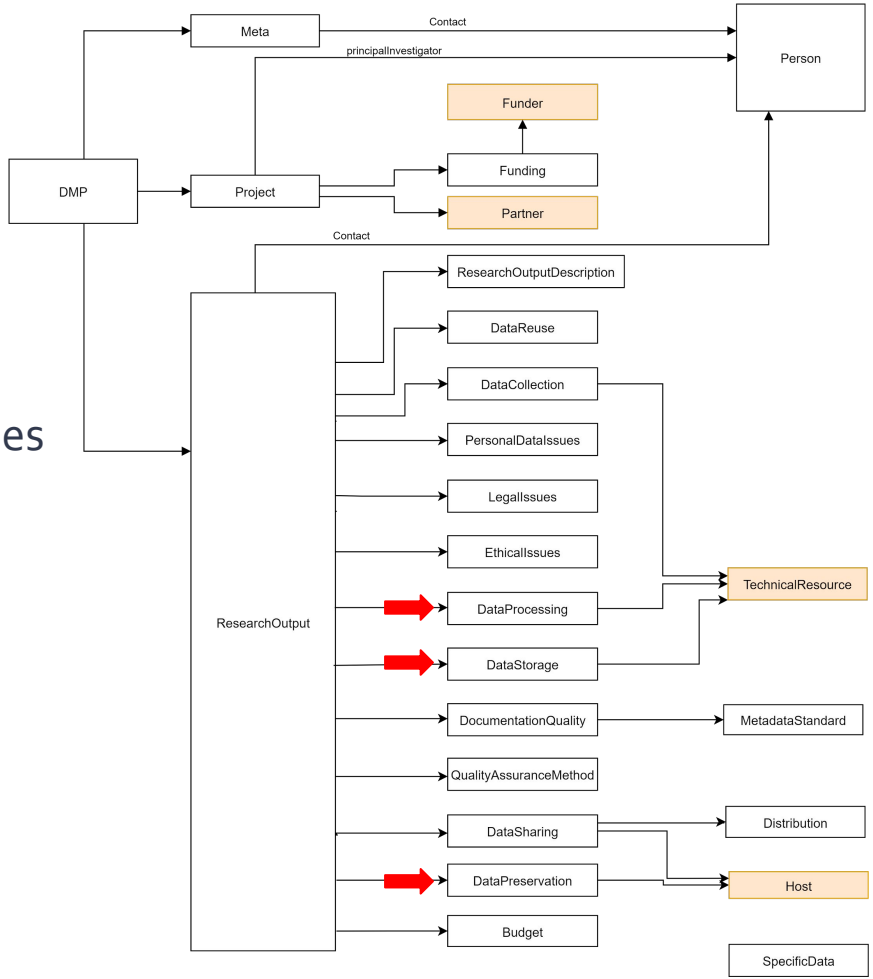


Model overview

Top-level entries :

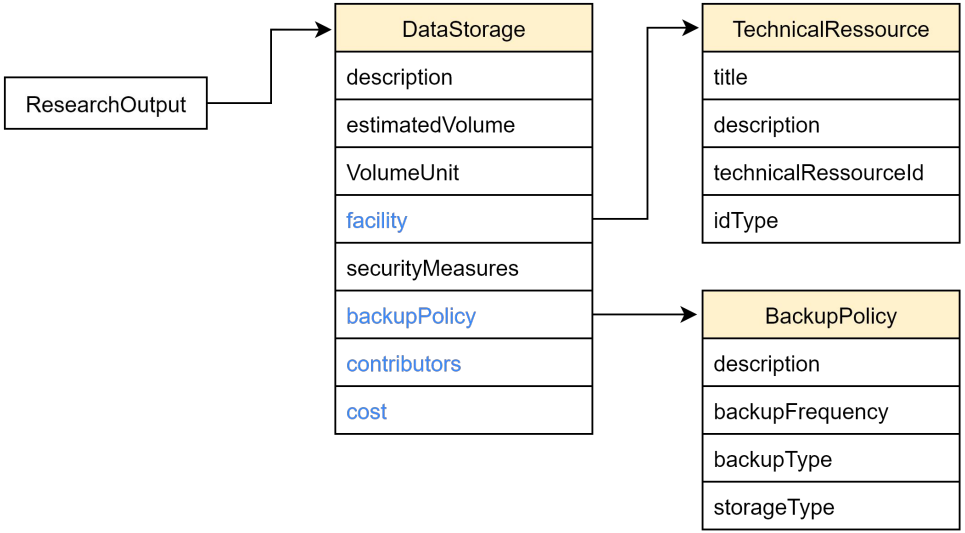
- Meta : metadata on the project
- Project : info about the project
- researchOutput : one or several outcomes of the project

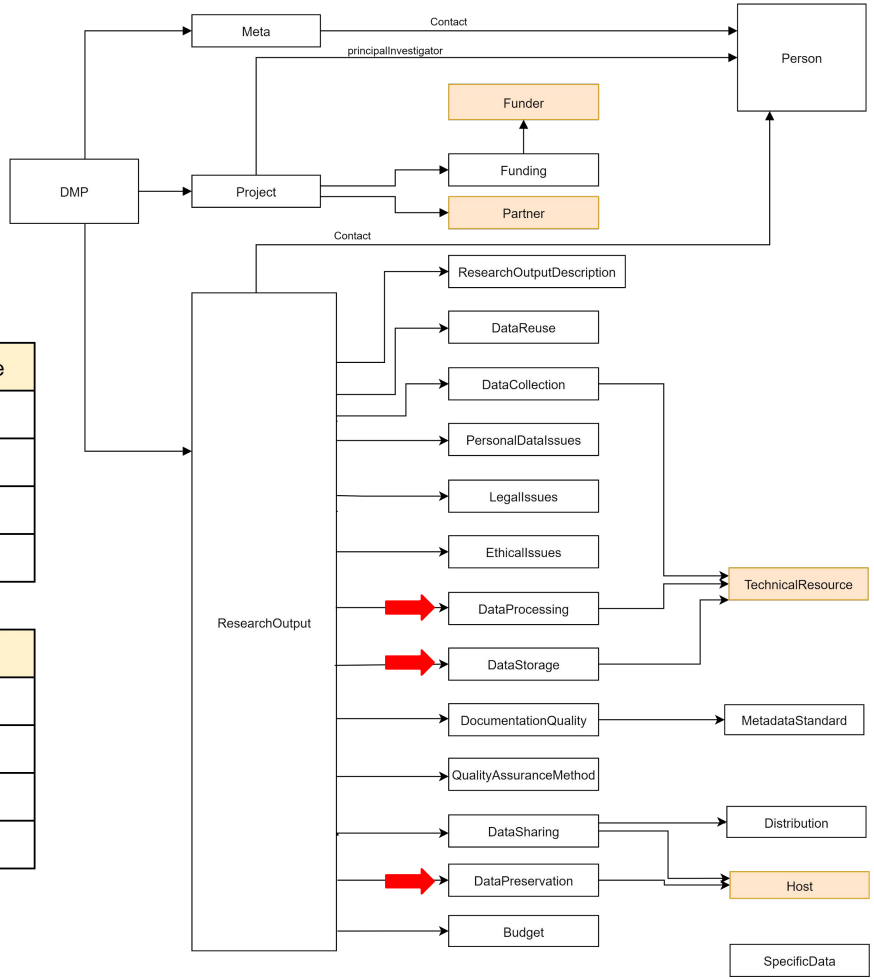




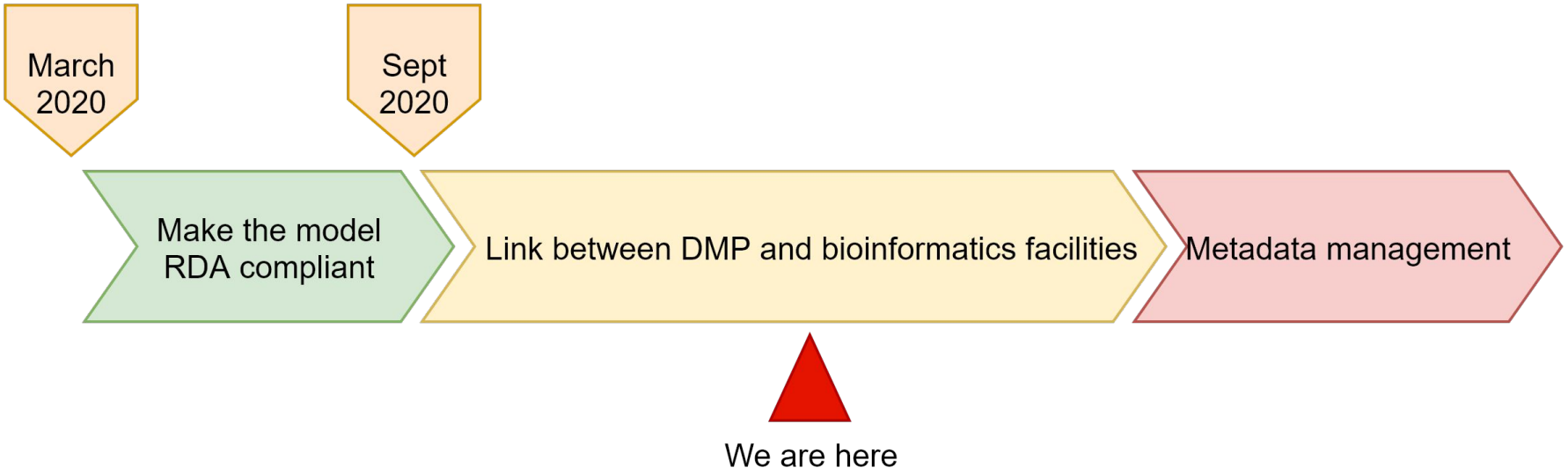
Model overview

Deeper levels of structuration



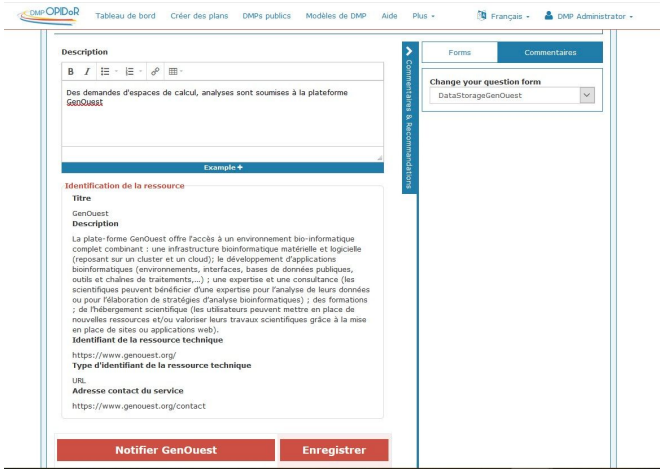


Project structuration



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



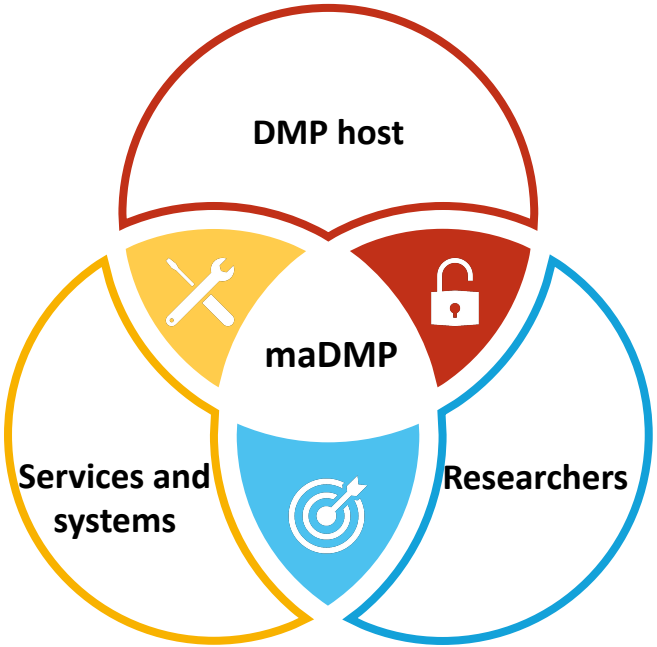
maDMP for researchers

- From an administrative chore to an asset
  - Fully integrated in the research process

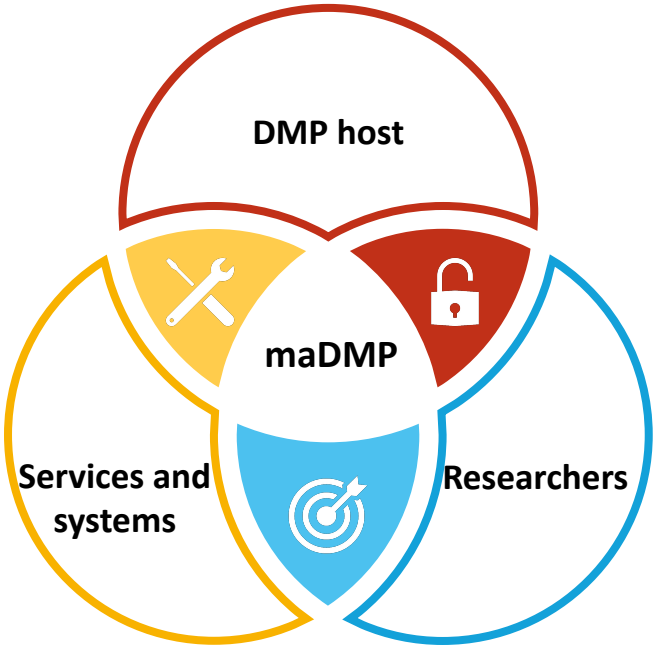
- Associated to a Digital Object Identifier
  - Traceability of both data and processes involved in the project

- 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



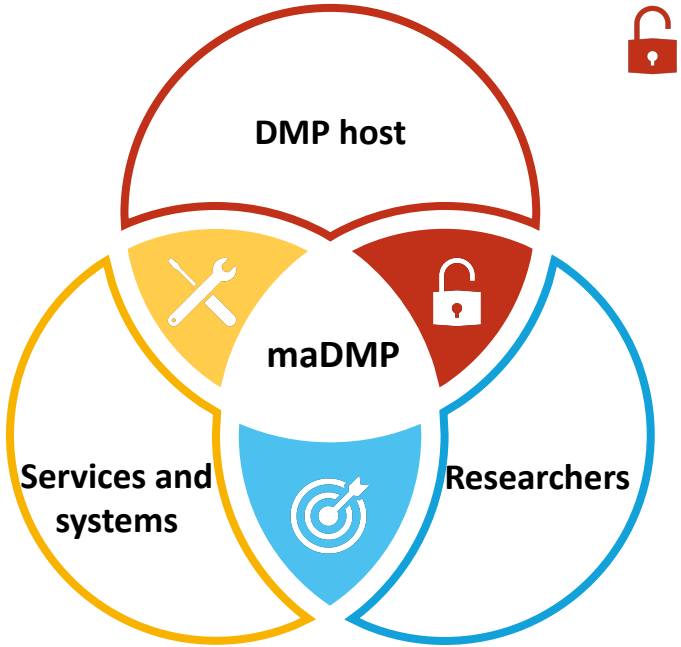
The benefits of a machine actionable DMP



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



FAIR access

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

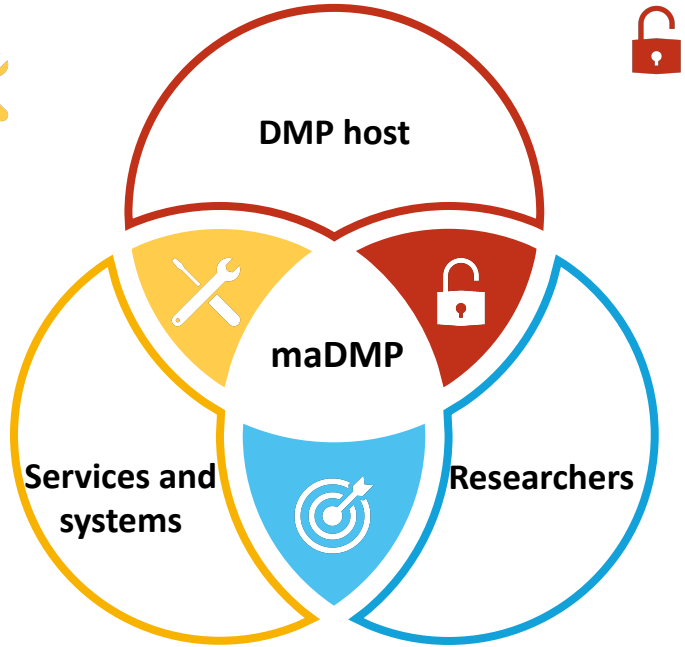
Communication

- 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

Data Management

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



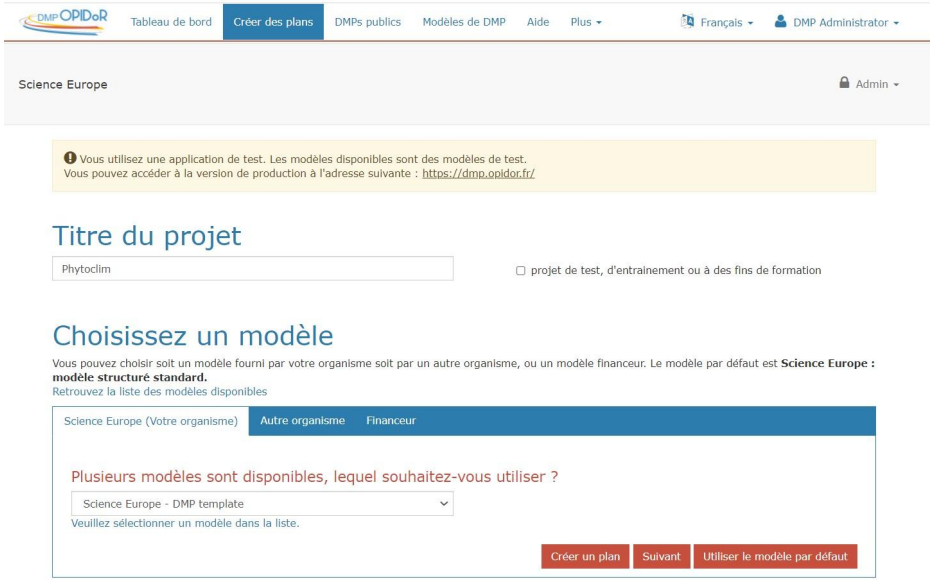
FAIR access

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

Communication

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

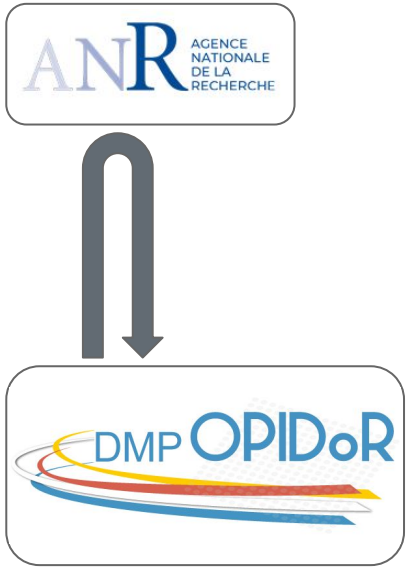
The future data life cycle



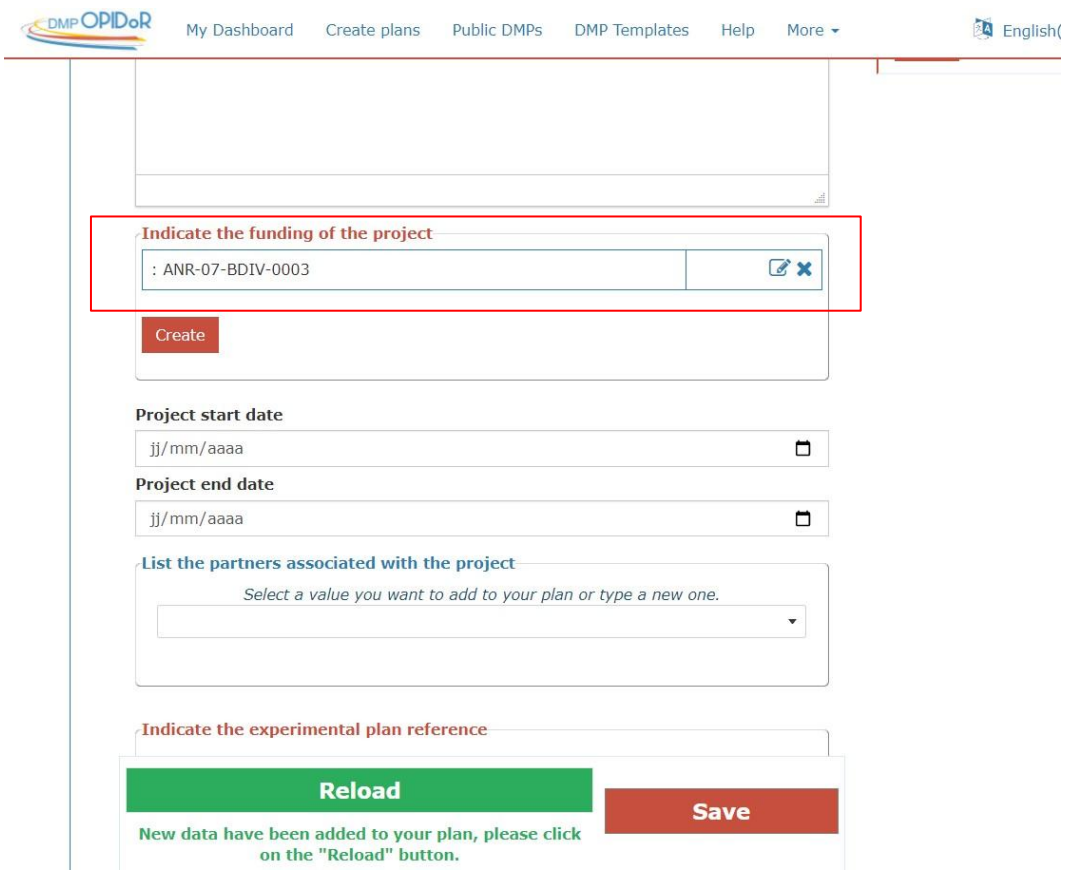
1. **Start writing the DMP**

The future data life cycle

2. Get the project informations

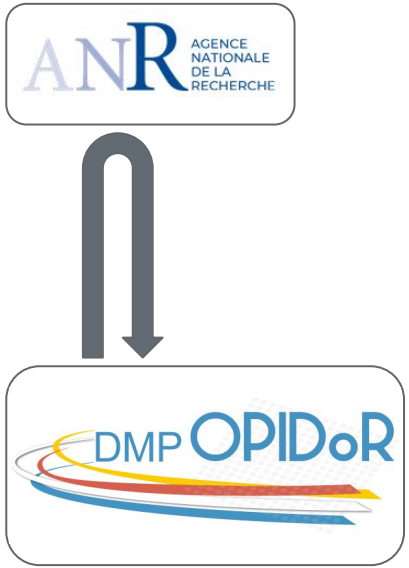


1. Start writing the DMP

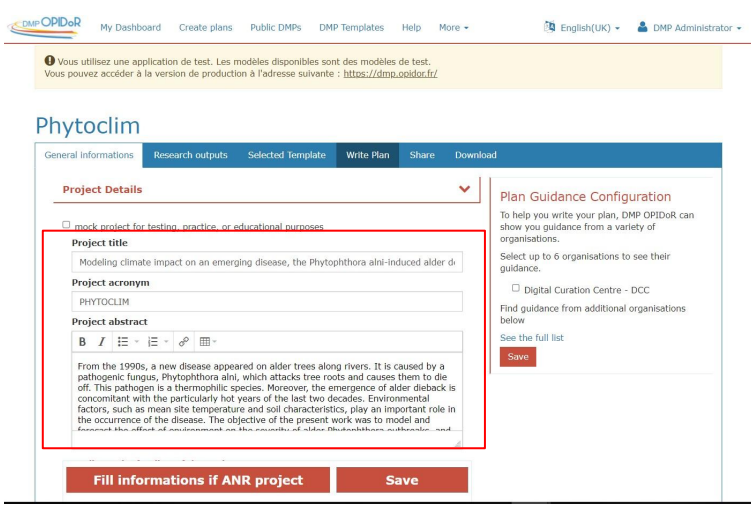


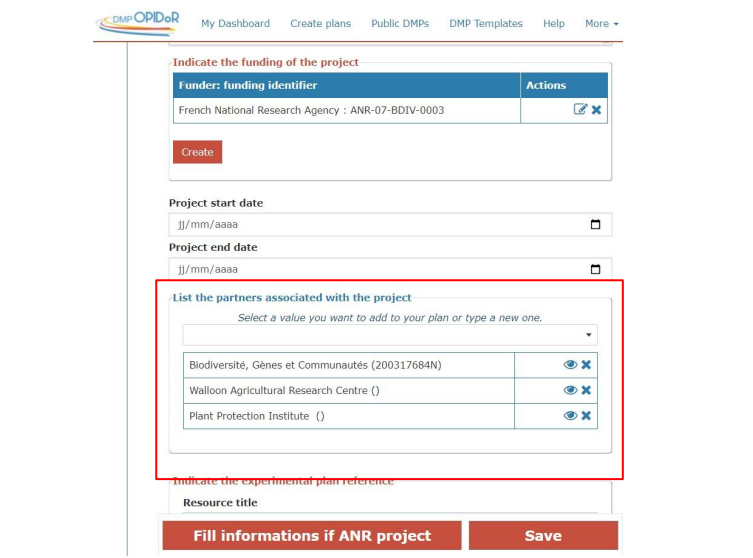
The future data life cycle

2. Get the project informations



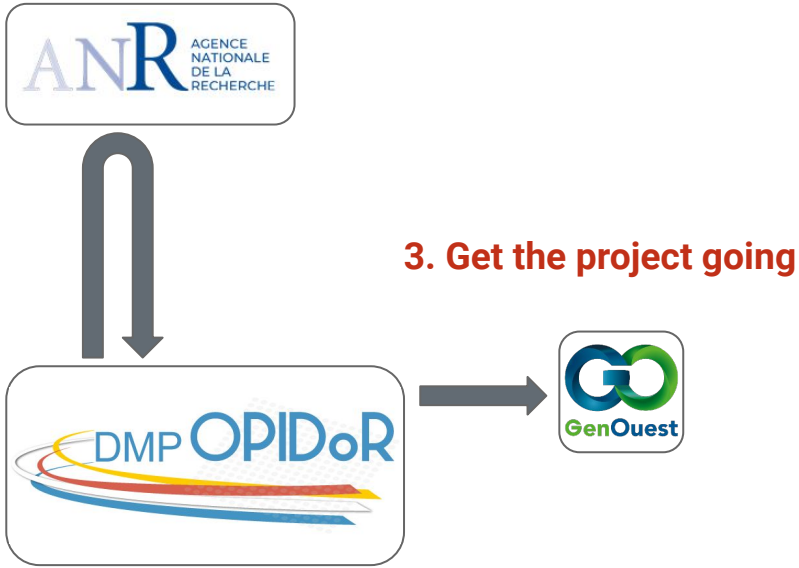
1. Start writing the DMP





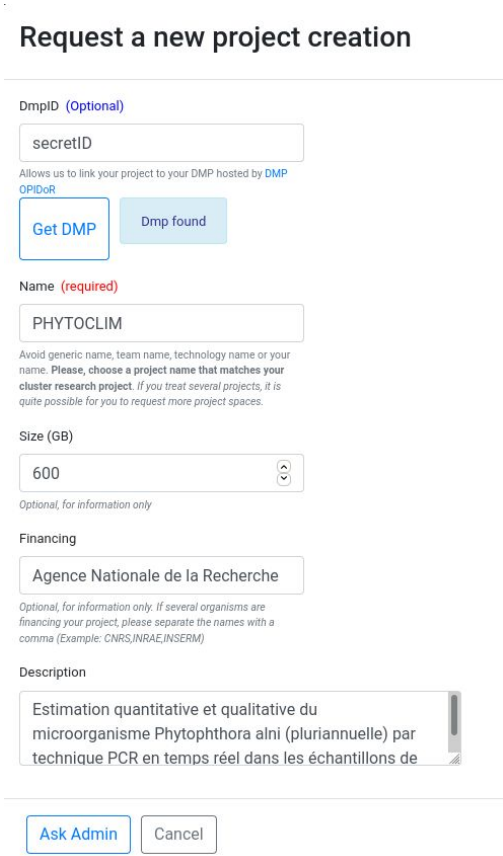
The future data life cycle

2. Get the project informations



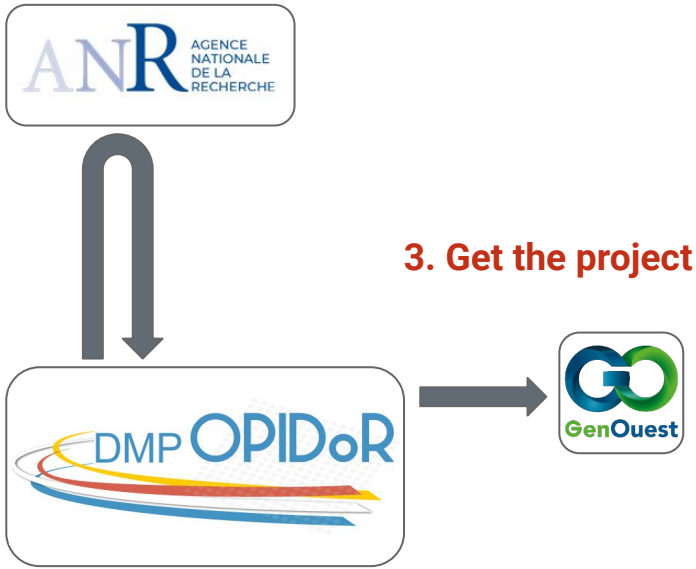
3. Get the project going

1. Start writing the DMP

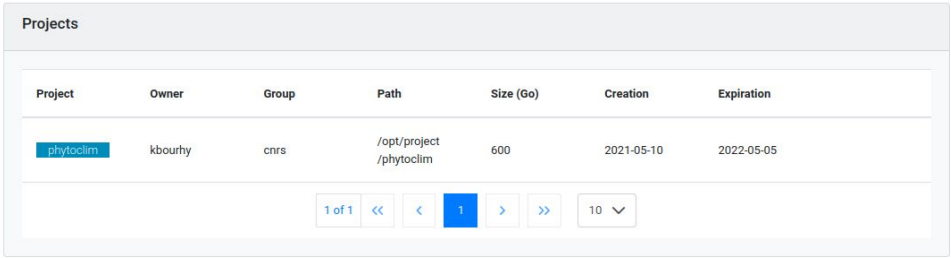


The future data life cycle

2. Get the project informations



3. Get the project going



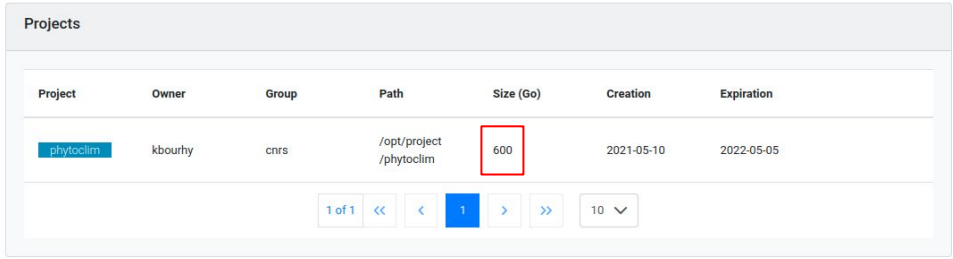
- Workspace specificities automatically established
- Collaborators with a Genouest account added

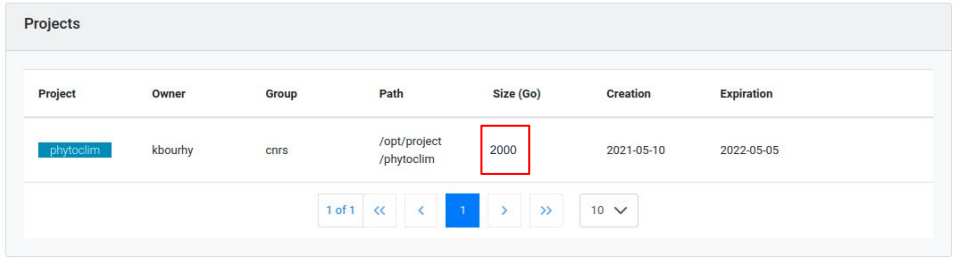
1. Start writing the DMP

The future data life cycle

2. Get the project informations

3. Get the project going





1. Start writing the DMP

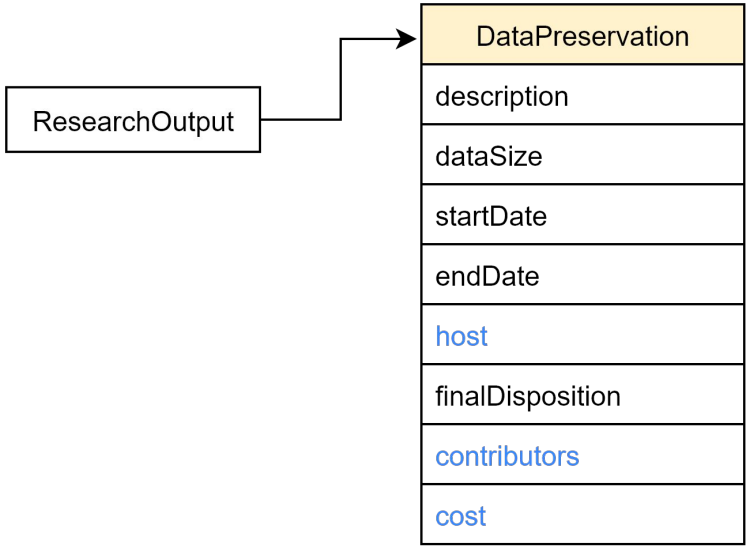
The future data life cycle

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/