maDMP Elixir-FR

IFB - Inist

Olivier Collin, Marie-Christine Jacquemot

elixir All Hands Meeting 2020

elixir-europe.org

maDMP4LS

Mid 2019 : ANR call Open Science : research practices and open research data

The objective of this call was to accelerate the adoption of practices for accessibility, reuse, openness of research data.

Consortium between IFB and Inist proposed “machine actionable DMP for Life Sciences”

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

Consortium

IFB : Institut Français de Bioinformatique / French Bioinformatics Institute

- National Network of Computing resources (NNCR)
- Core-cluster + Core-cloud + 21 bioinformatics facilities in France
- Elixir-FR node

Inist : Institut National de l'Information Scientifique et Technique

- Provider of OPIDoR tools (Optimiser le Partage et l'Interopérabilité des Données de la Recherche)
  - DMP-OPIDoR : planning
  - Cat OPIDoR : identifying services
  - PID OPIDoR : DOI service

DMP : 3367
Templates : 18
Users : 3911

Project structuration

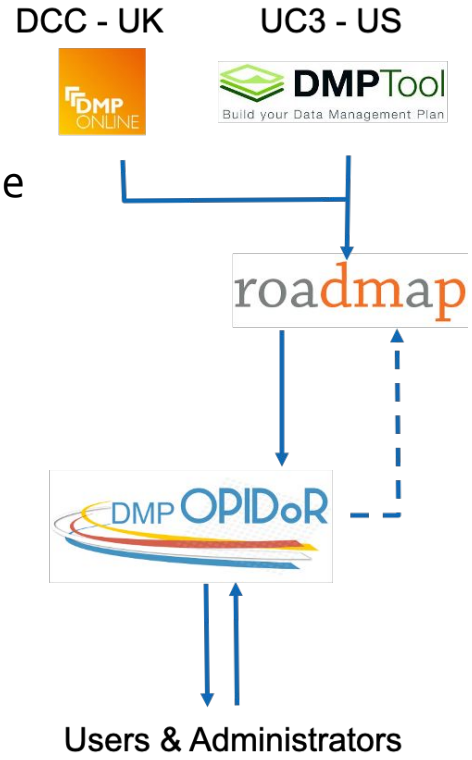
- WP1 : Interfacing the data analysis services provided by IFB with DMP-OPIDoR
  - Data model compliant with RDA DMP common model
  - Link between DMP and User/Project management on bioinformatics facilities
  - Metadata capture within SEEK
- WP2 : Fostering the adoption of new practices by the communities
  - Train the trainer actions
  - Regional training courses
- WP3 : Use cases and communities
  - EMBRC image
  - Links with communities

DMP OPIDoR evolution v1 & v2

- 2016 : launch of DMP OPIDoR based on DMPonline open source code
- June 2018 : migration based on DMP Roadmap common code

Adaptation to meet the French community needs

- Software developments
- Ergonomy : facilitate the choice of template
- Edition features
- Compliance with GDPR
- Multiple research products
- Resources, examples



From DMP OPIDoR to maDMP OPIDOR

- Maintain pedagogical and editorial features
- Produce a structured and standardized DMP content
- Use of internal/external registries and information systems :
  - to pre-populate DMP
  - to guide users through the selection of standards, or repositories, tools, etc. (FAIR principles)
- Enable exchange of informations with services and systems throughout the data life cycle
- Enable exchange of DMP content between different DMP tools using RDA maDMP application profile

From DMP OPIDoR to maDMP OPIDOR

- Adopt RDA recommendations

Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLoS computational biology*, *15*(3), e1006750. https://doi.org/10.1371/journal.pcbi.1006750

RDA RESEARCH DATA ALLIANCE

elixir CONVERGE elixir All Hands Meeting 2020

DMP-OPIDoR data model evolution

Methodology

Take into account:

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

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

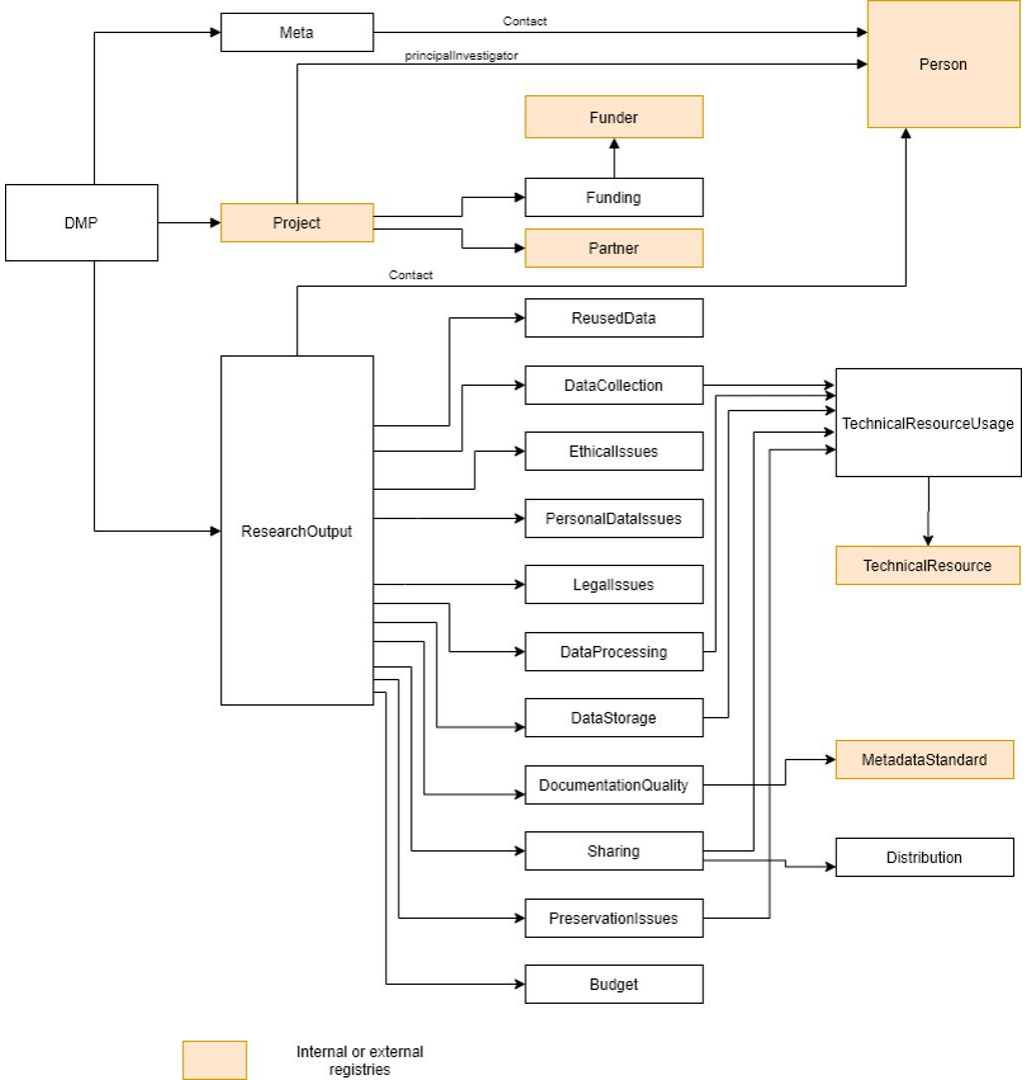
Output

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

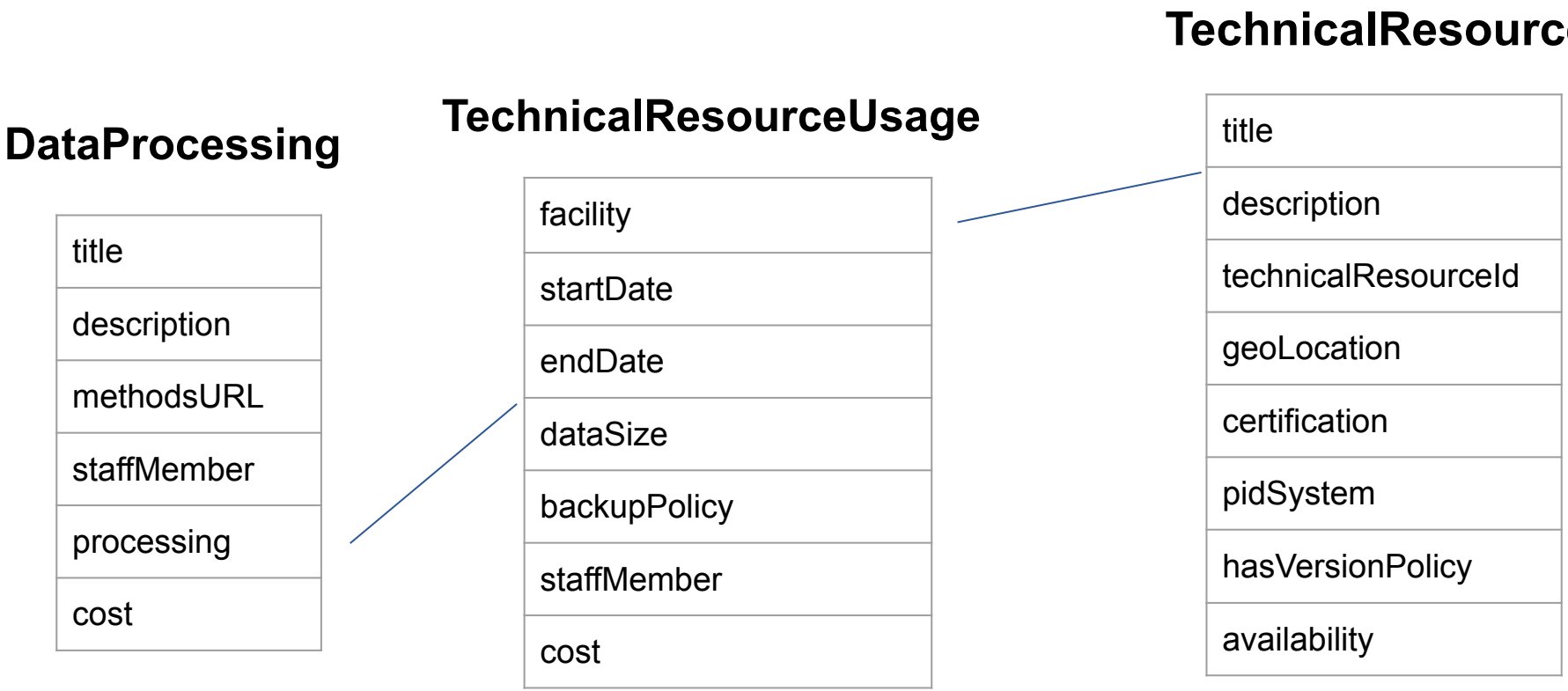
Model overview

Top-level entries :

- Identifier
- Meta : metadata on the project
- Project : info about the project
- researchOutput



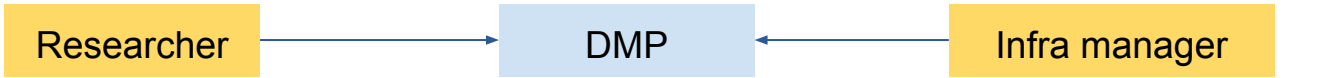
Deeper levels of structuration: data processing example

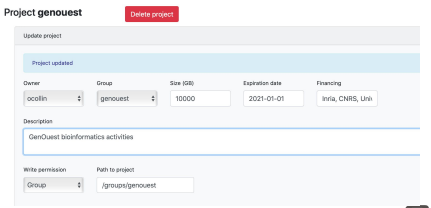


Increase efficiency in the management of hosted scientific data

- How much storage space should be provisioned for a project ?
- For how long ?
- What becomes of the data ?
- Which users ?
- Data access ?

From the infrastructure manager point of view

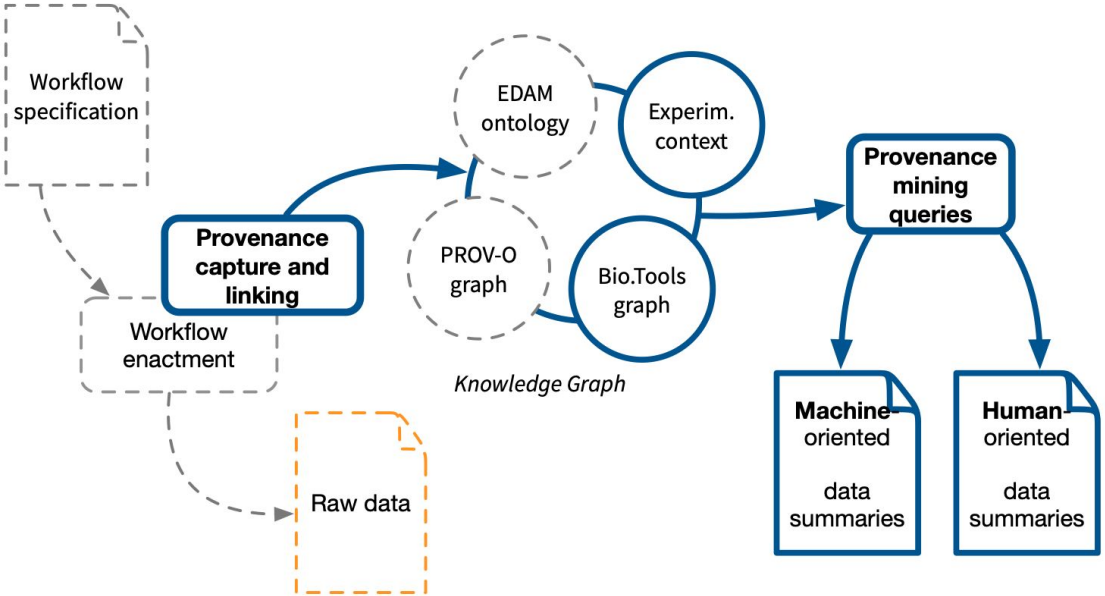




myIFB

Planned developments

Capturing the runtime information



Alban Gaignard et al. Findable and reusable workflow data products: A genomic workflow case study http://dx.doi.org/10.3233/SW-200374

Planned developments : IFB infrastructure project

MuDiS4LS : Mutualised Digital Space for Life Sciences
IFB answer to a national infrastructure call.

A framework relying on the national and regional data centers and mesocenters to enable scientists to orchestrate the fluxes of biological data :

- From the source (data-producing national infrastructures)
- Public release via national or international repositories
- Mid-term securing during the intermediate phases of analysis and exploitation.

maDMP4LS is a cornerstone

Thank you

Acknowledgments :
Inist OPIDoR team : J.-M. Parret, M.-C. Jacquemot, F. Mazur, B. Faure, F. Cosserat, A. Busin, L. Rassinoux.
IFB : J. van Helden, P. Lieby, A. Gaignard, NNCR Workgroup.

ELIXIR-Converge has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 871075.

@ELIXIREurope

/company/elixir-europe