



## DMP Common Standards WG

# RDA DMP Common Standard for machine-actionable Data Management Plans

Paulette Lieby [0000-0002-9289-9652](tel:0000-0002-9289-9652)



INSTITUT FRANÇAIS DE BIOINFORMATIQUE



INRAE



Inserm



BY

Chair (s): Marek Suchánek, Peter Neish, Tomasz Miksa

→ Essentially, develop a common data model with a core set of elements

→ Output (Dec. 2019): **RDA DMP Common Standard for Machine-actionable Data Management Plans**

◆ Github:

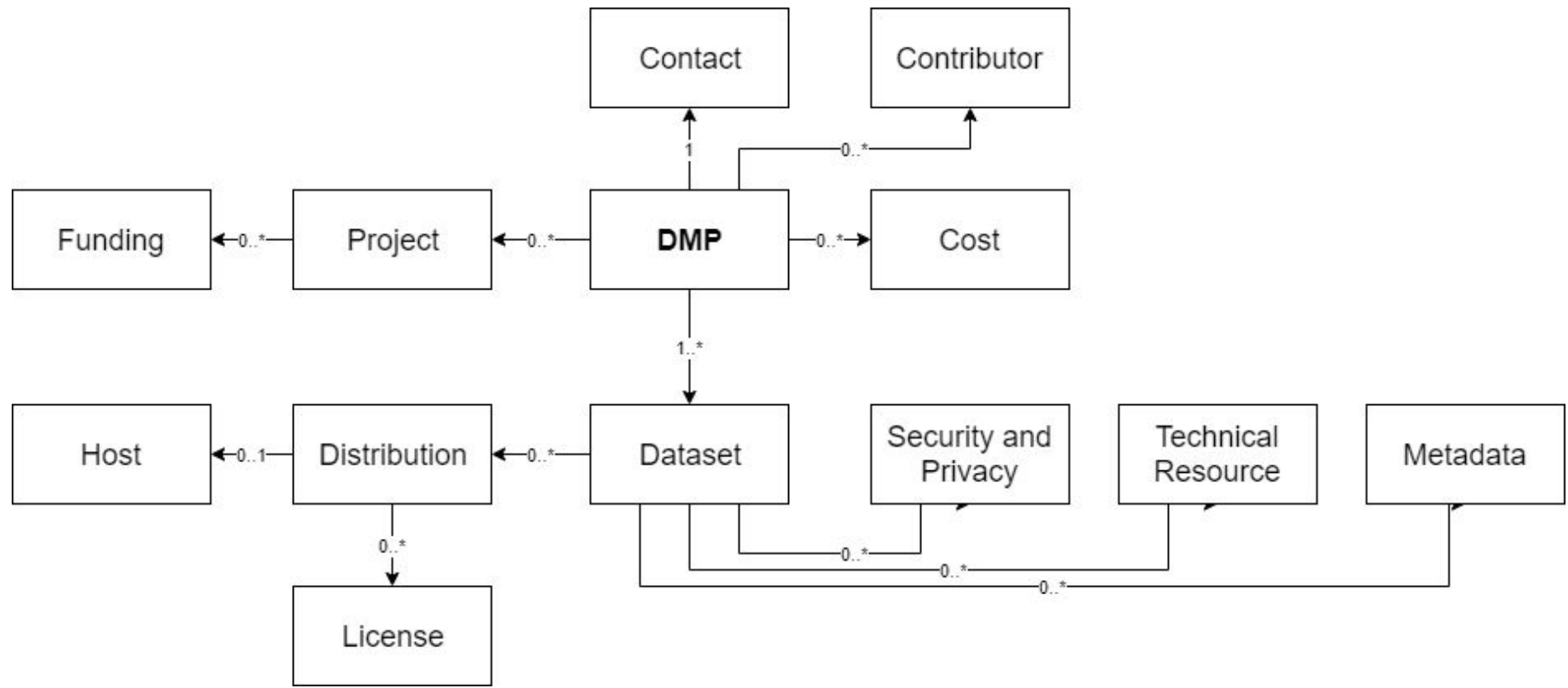
<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>

- A document that guides the planning and implementation of RDM
  - Application of good data management practice over the data lifecycle (data collection and creation, data processing, data sharing and preserving)
  - Type and volume of data, data organisation and quality, metadata, storage and backup, archiving, legal and ethical issues
- A DMP model is a framework that provides guidance in good data management practices
  - Two main models
    - Horizon 2020 (model follows the FAIR principles)
    - Science Europe (model follows the data lifecycle)
  - Science Europe is the most commonly used, and recommended by ANR (for French-funded research)

- A DMP as static text
  - Has no usable data (by machines)
    - Unless via text mining
  - Cannot collect pre-filled information from information systems
  - Might quickly get obsolete
- Such a DMP cannot be integrated as an ‘active’ part into the wider information system

- To allow automated information exchange between all stakeholders involved in Research Data Management (RDM)
  - Machine action on FAIR digital objects
    - A digital object is contextualised data that is readable by machines
  - **Autonomous** machine action via the DMP
- DMPs as evolving/active documents
  - To enable automated integration and updates
  - To enable monitoring, information exchange, validation

- A standard for automated information exchange between a DMP and other information systems, and between DMPs
- See [RDA DMP Common Standard for machine-actionable Data Management Plans](#) which describes the metadata application profile
- Relatively simple, a few classes
- The ‘hour glass’ principle
  - *“where “convergence” is driven by absolute minimal standards on the most generic operations, leaving maximum freedom to implement solutions in specialized domains and in use cases involving legacy data”* ([FAIR Digital Object Framework](#))



[RDA DMP Common Standard for Machine-actionable Data Management Plans | Zenodo](#)

[Application Profile for Machine-Actionable Data Management Plans](#)

## Properties in 'dmp'

Name	Description	Data Type	Cardinality	Example Value
<a href="#">contact</a>	Contact person for a DMP	Nested Data Structure	1	
<a href="#">contributor</a>	To list people that play role in data management related to this DMP, e.g. responsible for performing actions described in this DMP.	Nested Data Structure	0..n	
<a href="#">cost</a>	To list costs related to data management. Providing multiple instances of a 'Cost' allows to break down costs into details. Providing one 'Cost' instance allows to provide one aggregated sum.	Nested Data Structure	0..n	



## Towards semantic representation of machine-actionable Data Management Plans

- DCSO: DMP Common Standard Ontology
- DCSO core: identifiers classes and external classes
  - External classes are reused from imported ontologies:
    - W3 DCAT, FOAF, DCSX, DCMi
    - DCSX: DCSO extensions

- From DMP-OPIDoR
- DMP-OPIDoR's data dictionary is RDA CS-compatible
- But it is also richer

```
"dmp": {  
  "created": "2022-01-14T00:00:00",  
  "description": null,  
  "dmp_id": {  
    "identifier": "https://dmp.opidor.fr/plans/12209",  
    "type": "url"  
  },  
  "language": "fra",  
  "modified": "2022-01-14T00:00:00",  
  "title": "DMP du projet \"Paulette's Plan\"",  
  "contact": {  
    "contact_id": {  
      "identifier": null,  
      "type": null  
    },  
    "nbox": "paulette.lieby@france-bioinformatique.fr",  
    "name": "Lieby Paulette"  
  },  
  "contributor": [  
    {  
      "name": "Lieby Paulette",  
      "nbox": "paulette.lieby@france-bioinformatique.fr",  
      "role": [  
        "Coordinateur du projet",  
        "Personne contact pour les données (RO 1, RO 2)",  
        "Responsable du plan"  
      ],  
      "contributor_id": {  
        "identifier": null,  
        "type": null  
      }  
    }  
  ],  
  "cost": [  
    {
```

- From DSW

→ Allows for  
DMP-OPIDoR -  
DSW  
interoperability

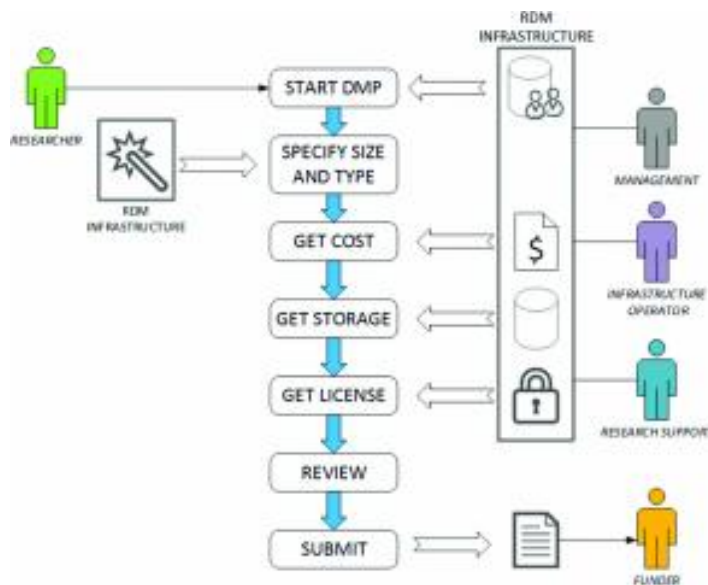
```
"dmp": {  
  "contact": {  
    "contact_id": {  
      "identifier": null,  
      "type": null  
    },  
    "nbox": "guillaume.gay@france-bio-imaging.org",  
    "name": "gay guillaume"  
  },  
  "contributor": [  
    {  
      "contributor_id": {  
        "identifier": null,  
        "type": null  
      },  
      "nbox": "marc.mongy@france-bioimaging.org",  
      "name": "mongy marc",  
      "role": [  
        "Responsable curation des donn\u00e9es"  
      ]  
    },  
    {  
      "contributor_id": {  
        "identifier": "0000-0002-9642-7994",  
        "type": "ORCID id"  
      },  
      "nbox": "edouard.bertrand@france-bioimaging.org",  
      "name": "bertrand edouard",  
      "role": [  
        "Responsable scientifique"  
      ]  
    }  
  ]  
}
```

- All major DMP tools
  - DMP Online by Digital Curation Centre (DCC) in the UK
  - DMP Tool by California Digital Library (CDL) in the US
  - DMP OPIDoR by Centre national de la recherche scientifique (CNRS) in France
  - RDMO by Leibniz - Institut für Astrophysik Potsdam in Germany
  - Data Stewardship Wizard (DSW) by Elixir research infrastructure in the EU
  - Argos - OpenDMP by OpenAIRE and EUDAT research infrastructures in the EU
  - F1000Research - open research publisher in the UK
  - easyDMP in Norway
  - DAMAP by TU Wien, TU Graz, and University of Vienna

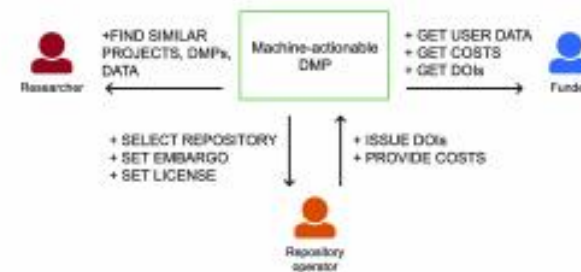
- Possibly an extension since the existing data dictionary is rather poor (of course, because it is simple)
- But extensions are not a given (because of the point above: simplicity)
- And cannot be done within the existing RDA WG (which is in maintenance mode)

## Automating Research Data Management Using Machine-Actionable Data Management Plans

- Machine-actionable data management planning in an institutional context
- Leveraging automated data flows allowed by maDMPs

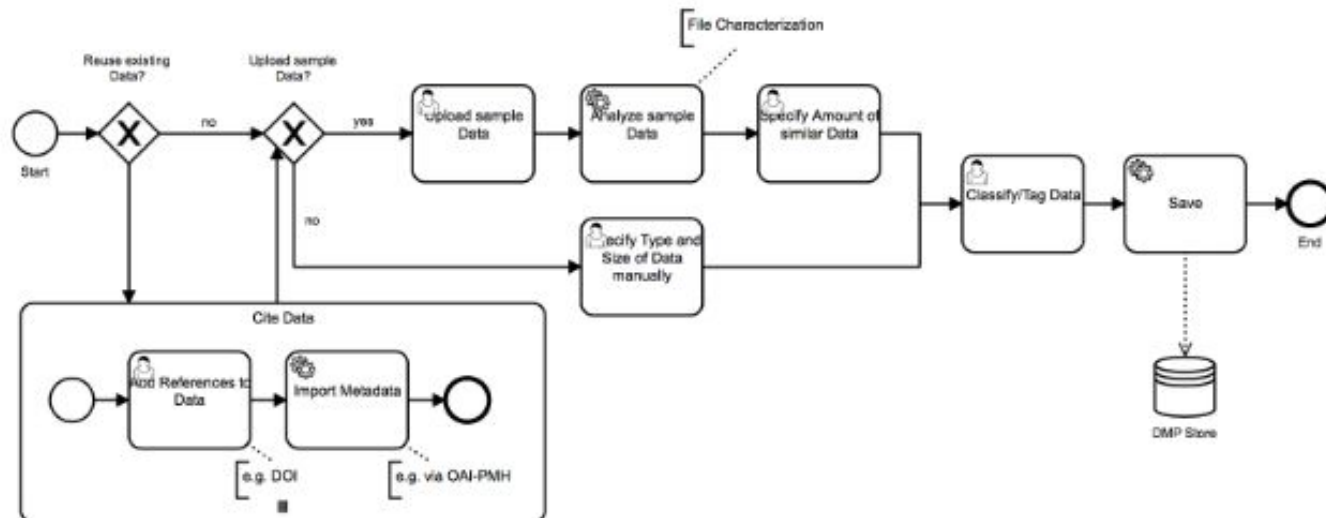


(a) Initial project phase [MNWR18]



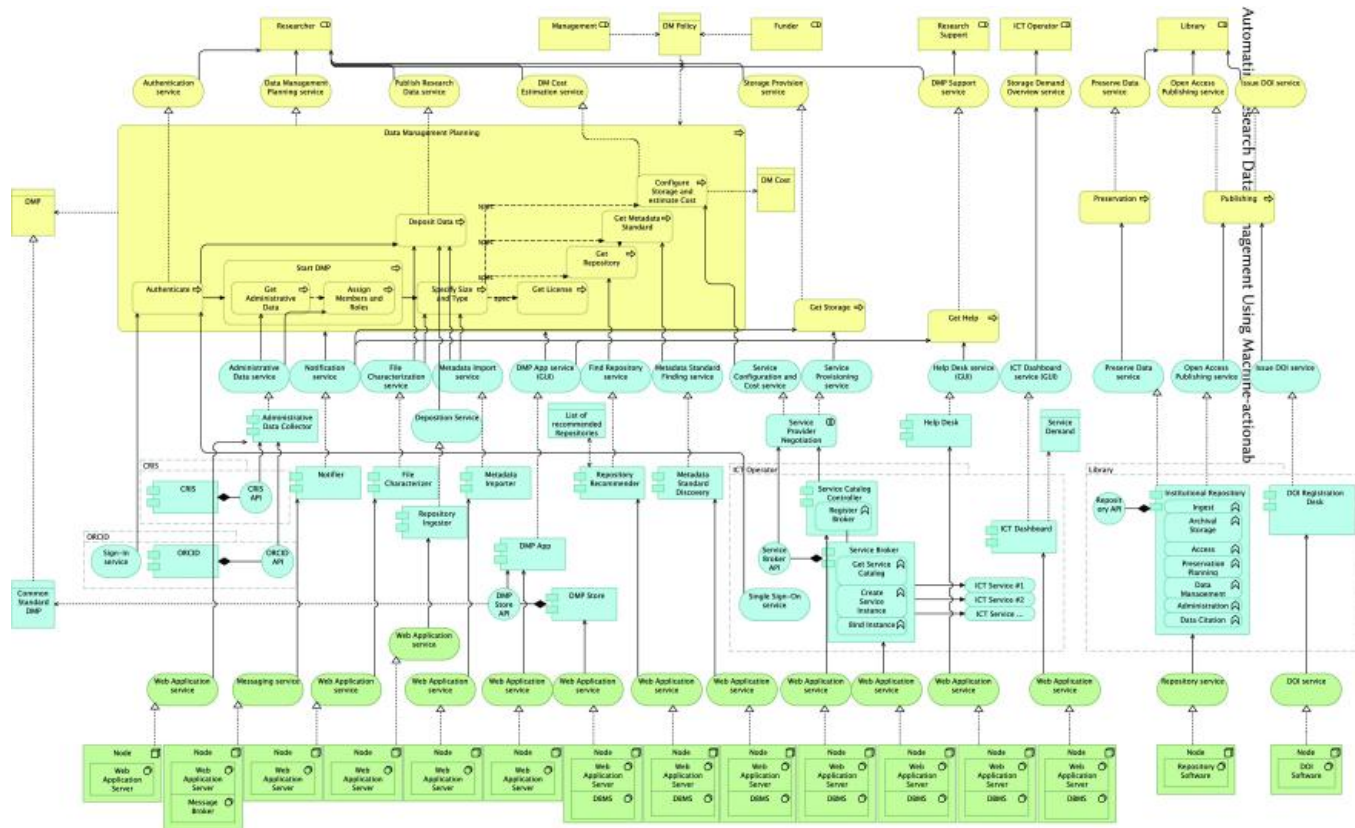
(b) Later project phase [MSMJ19]

- Specify data size and type workflow



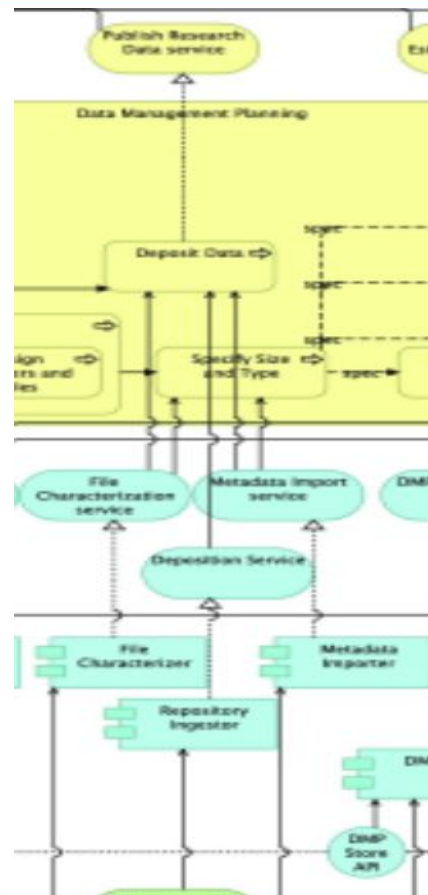
- Full architecture layers

- Business
- Application
- Technology





- Specify data size and type workflow
  - Business
  - Application



- [list of papers](#)

Questions?



INSTITUT FRANÇAIS DE BIOINFORMATIQUE

