Turning Machine-actionable DMPs into FAIR Digital Objects

Tomasz Miksa

TU Wien and SBA Research, Austria

tmiksa@sba-research.org

Informatics

Data Management Plans (DMPs)

	Data Officer	Who is responsible for the data management and the DMP of the project (name/email address)?
I	**Data Characteristics**	
I.1	**Description of the data**	What kinds of data/source code will be generated or reused (type, format, volume)? How will the research data be generated and which methods will be used? How will you structure the data and handle versioning? Who is the target audience?
II	**Documentation and Metadata**	
II.1	**Metadata standards**	What metadata standards (if any) will be in use and why? (see Digital Curation Centre)
II.2	**Documentation of data**	What information is needed for the data to be findable, accessible, interoperable and re-usable (FAIR) in the future? Is the data machine-readable? How are you planning to document this information?
II.3	**Data quality control**	What quality assurance processes will you adopt? How will the consistency and quality of data collection be controlled and documented? (This may include processes such as repeat samples or measurements, standardised data capture, peer review of data or representation with controlled vocabularies.)
III	**Data Availability and Storage**	
III.1	**Data sharing strategy**	How and when will the data be shared and made accessible? What repository will you be using? What persistent identifier will be used?
III.2	**Data storage strategy**	What data are to be preserved for the long-term, and what data will not be stored? How and where will the data be stored and backed up during the research? How and where will the data be stored after the project ends? For how long will the data be stored? Are there any costs that need to be covered for storage? At what point during or after the project will the data be stored? Are there any technical barriers to making the research data fully or partially accessible?

Directorate for Engineering
Data Management Plans Guidance for Principal Investigators
updated: November 2018

The Directorate for Engineering (ENG) supports research covering a broad spectrum of communities of investigators, and each community has its own best practices. ENG is aware of the need to provide flexibility to programs, principal investigators (PIs), and reviewers in assessing the quality of individual Data Management Plans (DMPs) from various communities. Therefore, guidance has evolved to accommodate changing community standards and expectations. ENG relies on the merit review process to determine which DMPs best serve each community.

The following guidance is to assist ENG investigators, reviewers and Program Officers in developing and evaluating effective, complete, and competitive DMPs. It is important to recognize that while all DMPs should address the five categories of information as specified in the PAPPG, they should not be generic. Each DMP should appropriately identify the data, metadata, samples, software, algorithms, curricula, documentation, publications, and other materials generated in the course of the proposed research. Moreover, the DMPs should describe how these materials will be disseminated, made accessible, and archived while incorporating the best practices and standards for the proposed research. DMPs are subject to peer review. Please contact your specific Program Officer if you have any questions related to DMPs in the program context.

PAPPG and NSF-WIDE REQUIREMENTS

All proposals must include a supplementary document of no more than two pages labeled "Data Management Plan," as described in PAPPG Chapter II.C.2.j. The DMP is NOT part of the 15-page Project Description. *Proposals that do not include a Data Management Plan will be returned without review.*

You may request funds to cover costs of publication, page charges, or preparation of data as a direct cost in your budget proposal, which is evaluated as part of the merit review process. Any costs associated with implementing the DMP should be explained in the Budget Justification.

Some NSF Program Solicitations may contain specific and/or additional instructions that deviate from this guidance and/or provide exceptions to the two-page limit. Instructions in the solicitation take precedence over this guidance. Please check solicitations carefully for this information.

DATA MANAGEMENT PLAN (DMP) CONTENT

The DMP should clearly articulate how the investigators plan to manage and disseminate both the physical and digital data generated by the project, taking advantage of emerging information

DMPs vs maDMPs

Traditional DMP

```
<administrative_data>
    <question>Who is responsible for the DMP?</question>
    <answer>Moritz from our university.</answer>
</administrative_data>
```

Machine-actionable DMP

```
"contributor" : [ {
  "contributor_id" : {
    "identifier" : "0000-0002-5164-2690",
    "type" : "orcid"
  },
  "mbox" : "moritz.staudinger@tuwien.ac.at",
  "name" : "Moritz Staudinger",
  "role" : [ "Data Manager" ]
```

RDA DMP Common Standard for Machine-actionable Data Management Plans

The Challenge:

Data Management Plans are free-form text documents describing the data that is used and produced during the course of research activities. They specify where the data will be archived, which licenses and constraints apply, and to whom credit should be given, etc. The workload and bureaucracy often associated with traditional DMPs can be reduced when they become machine-actionable.

Produced by: **DMP Common Standards WG**
https://www.rd-alliance.org/groups/dmp-common-standards-wg

Informatics

How tools implement it

- DMP tools import/export maDMPs

- machine-actionable DMPs are not yet truly living documents updated fully or partially by different stakeholders

Does turning maDMPs into FDOs and connecting them tighter with other FDOs can be the necessary next step to achieve better machine-actionability?
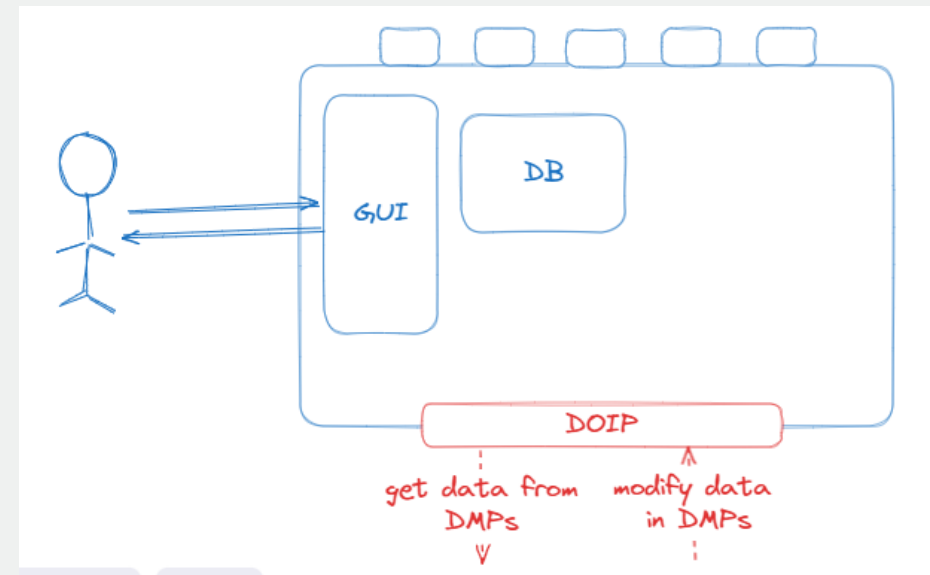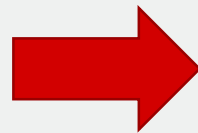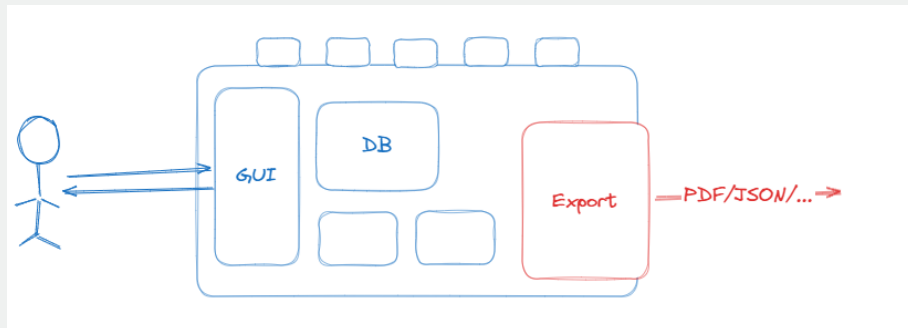
maDMP pointing to FDOs

- AS-IS
  - maDMPs consist of DCAT Datasets
  - Repeats basic information, e.g. license, location, host, etc.
- TO-BE
  - Use FDO
  - No repetition of information
    - Fetching directly from FDOs
- Consequence
  - DMP removed to the glue between Projects, People, FDOs
  - Provide human narrative
- Open question
  - What information can be taken for granted and which should still be in the DMP?
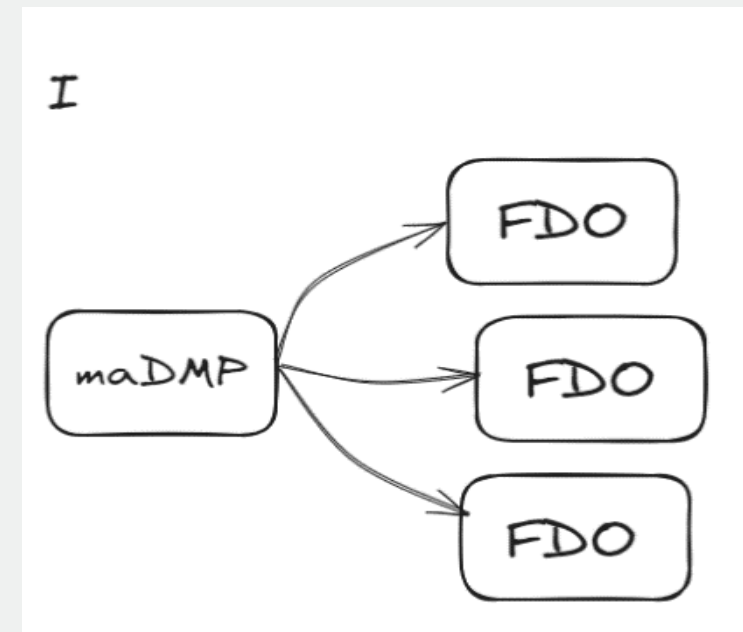
maDMP as an FDO

- maDMP can also be an FDO itself

- Dedicated type and a set of operations must be defined

- Operations
  - typical read-and-write operations to support the current use cases
    - e.g., to read out information on the associated project, location of data, access restrictions, etc.
  - More complex operations
    - e.g. the health check of the DMP

II

maDMP stores based on FDOs

- DMP tools -> maDMP Stores
- maDMP Store
  - repository system
  - registry system
- maDMPs have their own identifiers
- maDMPs can point to other FDOs
- Clients get to see different parts
  - depending on the operation permitted
- To be defined
  - own type for maDMPs
  - basic metadata and basic operations

TU WIEN Informatics
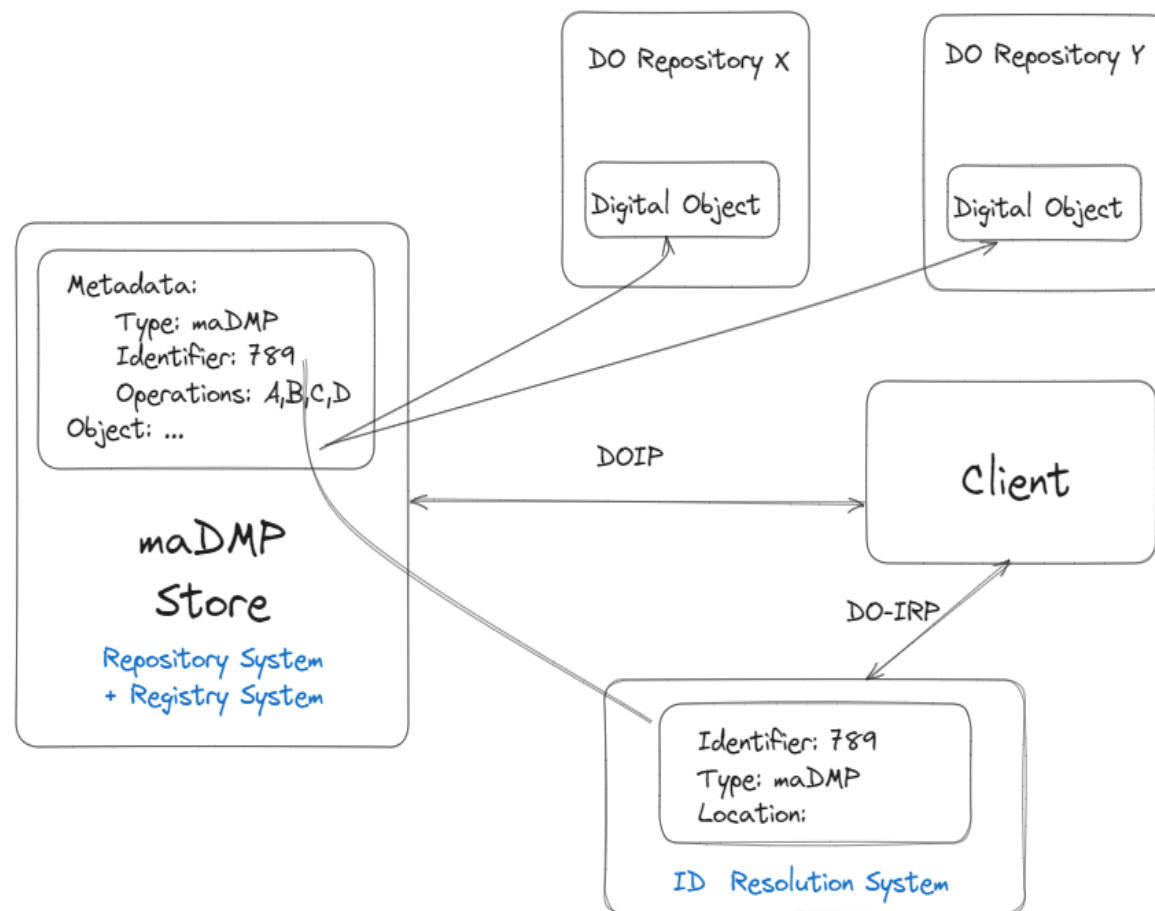
Summary

- Three different modifications are possible
  - maDMPs as FDO
  - maDMPs pointing to FDOs
  - DMP tools becoming maDMP Stores

- Developments needed
  - Own type for maDMPs
  - Basic metadata
  - Set of operations
  - ...

- Open questions
  - What information can we depend on?
  - Are maDMPs needed at all in the world consiting of FDOs?

https://ostrails.eu

Informatics

Publications about maDMPs

Tomasz Miksa, Marek Suchanek, Jan Slifka, Vojtech Knaisl, Fajar J. Ekaputra, Filip Kovacevic, Annisa Maulida Ningtyas, Alaa El-Ebshihy, and Robert Pergl. Towards a toolbox for automated assessment of machine-actionable data management plans. *Data Science Journal*, Aug 2023.

Tomasz Miksa, Simon Oblasser, and Andreas Rauber. **Automating research data management using machine-actionable data management plans**. ACM Transactions on Management Information Systems, 13(2), dec 2021.

Tomasz Miksa, Paul Walk, Peter Neish, Simon Oblasser, Hollydawn Murray, Tom Renner, Marie-Christine Jacquemot-Perbal, João Cardoso, Trond Kvamme, Maria Praetzellis, Marek Suchánek, Rob Hooft, Benjamin Faure, Hanne Moa, Adil Hasan, and Sarah Jones. **Application profile for machine-actionable data management plans**. CODATA Data Science Journal, 20(1):32, October 2021

Informatics