



Development project for maDMP deployment in Oulu University

08.04.2024

Anna Rohunen (anna.rohunen@oulu.fi)

Ivan Melobenskyi (ivan.melobenskyi@oulu.fi)



Starting points

- Goal: increasing a large organisation's research data management efficiency and making data management operations streamlined by means of digitalisation
- *Machine-actionable, dynamic data management plans (maDMP)* are employed in the system being developed
- Piloting in Information Technology and Electrical Engineering faculty (ITEE), results will be applied in all UO faculties



Key data management aspects and maDMP-based data management system (DMS) functions

Data management aspect	Examples of maDMP-based DMS functions
Data management plan (DMP) process	Drawing up the DMP: the DMS helps to create a DMP in line with the funder's requirements
Data management planning	Describing data: the researcher inputs information about the data to be collected or produced in a project
Data privacy	Guiding the researcher to process personal data in line with data protection legislation: e.g., initiating and conducting the risk assessment process if personal data are processed in a project
Communicating service needs to UO service providers	Communicating ICT service needs: informing ICT services about the researcher's data storage, processing, and sharing needs
Aggregate-level monitoring	Monitoring DMPs: keeping track of the number of draft and approved DMPs



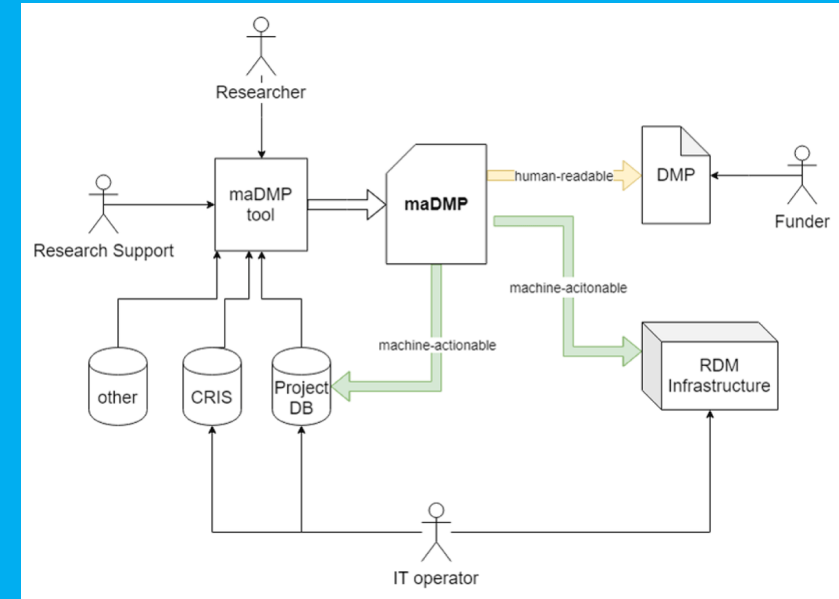
Examples of the stakeholders interaction in the DMS via data transfers

- Each **stakeholder** has his personal account and functionalities related to his needs. For example, ICT service personnel, library specialist, data steward and researchers.
- When a **researcher** is filling in a maDMP, for example, about storage needs, data archiving and personal data processing, relevant stakeholders listed above get an immediate notification about that and can take actions (for example, **ICT services** can allocate the needed storage, **data stewards** can check, if researchers' decisions are in compliance with GDPR and **library specialists** can help with the proper creation and publication of metadata.

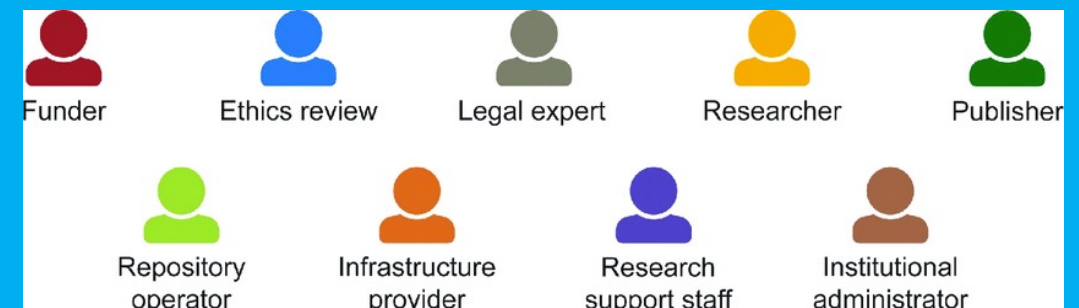


University of Oulu maDMP use cases

- Example A: Efficient communication between the *researchers* and *ICT services*
- Example B: FAIR evaluation with maDMP
- Example C: Support for identification of datasets for digital preservation service (PAS)
- Example D: Support for data protection processes
- Example E: Data protection reporting
- Other examples of the DMS functionalities based on maDMP information content



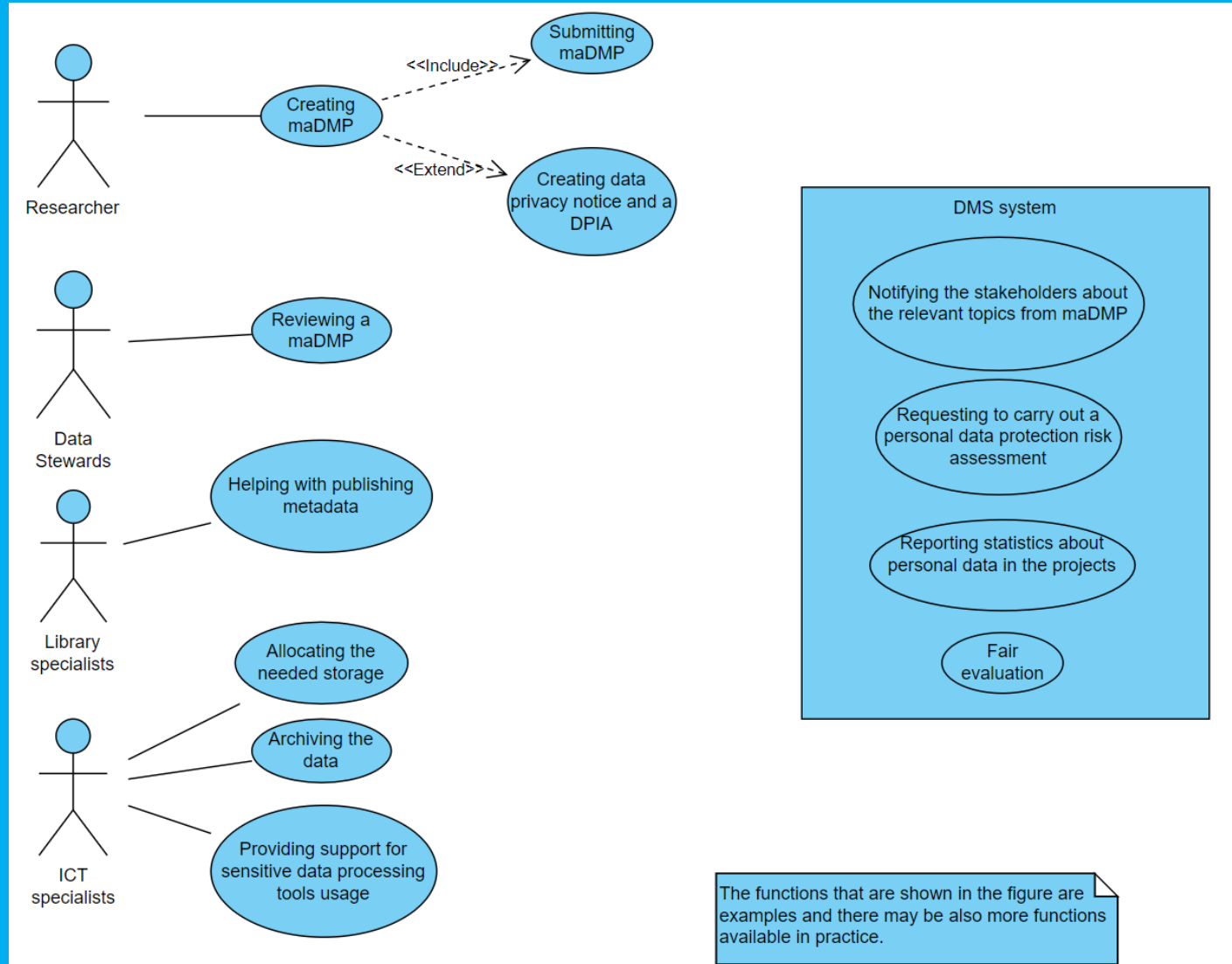
Source: Miksa T et al. (2021) Application Profile for Machine-Actionable Data Management Plans. Data Science Journal 20(1):32 DOI:10.5334/dsj-2021-032



Source: Miksa T, Simms S, Mietchen D, Jones S (2019) Ten principles for machine-actionable data management plans. PLoS Comput Biol 15(3): e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>



Use case diagram





Example A: Efficient communication between the researchers and ICT services

The **researcher** creates a maDMP in the data management system (DMS). The researcher realizes the need of data storage of 2 TB. When he makes the needed choice in maDMP, this information transfers to the **ICT services** that immediately take actions related to the needs of the researchers. ICT services allocate and map the needed storage drive and prepare it for the researcher's needs.

The **researcher** also needs a specific tool to process sensitive data. After choosing this option during a DMP creation, this information is transferred by **the DMS** to the **ICT services**. After they get informed about it, **a specialist** contacts the researcher and provides him support to process sensitive data with an appropriate tool.

The **researcher** needs to archive part of the data at the end of the project and dispose other parts of it. At the end of the project lifecycle, the researcher updates this information in the DMS. **ICT services** get notified just in time and provide with according assistance.



Example B: FAIR evaluation with maDMP

When a **researcher** is creating a maDMP, it is possible to notify him if his plans on data management correspond with FAIR principles. For example, **DSW** (Data Stewardship Wizard) has similar functionality, where FAIRness of the researcher's choice is shown by traffic light marks, so the researcher can adjust his choice and learn about data management options that comply with FAIR principles:

9 Will you need consent for any newly collected personal data? + !

a. No, We do not collect any new personal data
Reusability

b. No, our work on personal data can be done using another legal base ☰
Reusability

c. Yes, we will collect consent for our specific use of the data
Reusability

d. Yes, we will collect consent for our use as well as reuse of the data
Reusability



Example C: Support for identification of datasets for digital preservation service (PAS)

The **researcher** prepares a maDMP in the research data management system (DMS). The researcher has recognized significant re-use potential of the data to be collected in his study. The researcher selects 'yes' option for a question about whether their data will have significant re-use value.

The **system** recognizes the need for extra attention for metadata and documentation requirements and the need to agree on granting rights to the data for the university for long term preservation required by the PAS service. **PAS contact person** of the university is contacted so that the PAS process (acceptance process, metadata quality control etc.) can be started without delay when the data is ready.

In this case, the **researcher** is notified at an early stage about PAS requirements in order to collect high-quality metadata and agree on rights for using the PAS service beforehand.



Example D: Support for data protection processes

The *researcher* plans to collect personal data in his study. He selects ‘yes’ option in the maDMP template's question about personal data processing.

Once the DMP is ready and approved by the researcher, a data protection process is initiated by the *data management system (DMS)*. The researcher now automatically receives a request to carry out a personal data processing risk assessment, and is directed to fill out a risk assessment form in the DMS. This ensures that a personal data processing risk assessment is carried out before the processing starts (cf. GDPR requirements and Privacy by Design principles). The DMS also supports the researcher in other data protection tasks (e.g., creating a data privacy notice and a DPIA).

The *DMS* informs the data *protection steward* about the progress of the process. The data protection steward performs his tasks of the data protection process in the DMS (e.g., reviewing the risk assessment & the data privacy notice, initiating a DPIA). The data protection documentation is stored in the DMS, and access to this documentation is provided to the relevant parties (e.g., data protection steward, *data protection officer, management*).



Example E: Data protection reporting

The ***data protection officer (DPO)*** and ***data protection specialists*** need quantitative data on personal data processing activities and documentation for monitoring and planning purposes.

The ***data management system (DMS)*** has recorded maDMP information on the research projects that process personal data. The ***DPO*** can now retrieve, for example:

- Number of research projects that process personal data
- Number of research projects that process special categories of personal data
- Number of needed DPIAs



Other examples of the DMS functionalities based on maDMP information content

- Connecting the researcher to the support services that are needed in a specific project
- Informing the researcher about trainings based on their data types and research methods
- Communicating access and sharing policies to relevant stakeholders
- Transferring metadata to a metadata tool (e.g., Qvain, OuluCRIS)
- Monitoring the number of published metadata (reminders when needed)
- Checking validity of the dataset publishing plan (Right to open the data? Lawful basis for opening personal data? Are there sensitive personal data? Are the data documented well enough for reuse?)
- Searching for DMPs by a research unit, faculty, DMP name, keywords, creation date, funder, research infrastructure, ...
- Filtering DMPs by their status (draft, ready for a review, approved)
- Collecting aggregate information on researchers' needs for disk space, applications etc.