IMPLEMENTATION STORIES

FAIRSFAIR
Fostering Fair Data Practices in Europe

IMPLEMENTATION STORIES

Leveraging machine-actionable DMPs to build RDM infrastructure

Authors
Angus Whyte and Patricia Herterich
(Digital Curation Centre, University of Edinburgh)

From interviews with
Tomasz Miksa (TU Wien)

TU WIEN TECHNISCHE UNIVERSITÄT WIEN

Introduction

Data management plans (DMPs) can play a key role in helping researchers plan for FAIR data and support the creation of FAIR data. By implementing a DMP support tool that helps researchers write DMPs and then integrating this tool with institutional services that respond to the needs identified in the DMP, the DMP can become a more useful resource for producing FAIR digital objects and especially for ensuring these are properly stored and made accessible.

This implementation story provides an insight into creating a machine-actionable DMP tool, DAMAP, at the TU Wien, Austria and integrating the tool with other institutional services, beginning with data storage provision. This effort has gained support from FAIR Data Austria, a project running from January 2020 to December 2022 to strengthen knowledge exchange between Austrian universities and develop training, tools and support services to manage data in accordance with the FAIR principles throughout the whole research lifecycle. (Blumesberger et al., 2021)

FAIRsFAIR recommendation

"Institutions and research communities should prepare a roadmap and guidance for implementing machine-actionable DMP workflows"

FAIRsFAIR Recommendations on practice to support FAIR principles

Leveraging Machine-Actionable DMPs to Build RDM Infrastructure

Taking advantage of a new standard

Research funders now routinely expect their grant holders to prepare a data management plan (DMP). In recent years, providers of tools for researchers to write DMPs have been discussing 'maDMP' (machine-actionable-Data Management Plans). This term is used to express various ideas for making DMPs more actionable by services that could potentially respond in more automated ways to the needs that researchers have expressed in their project planning, including the steps that need to be taken to make data FAIR (Simms et al, 2017).

TU Wien senior researcher Dr Tomasz Miksa has been one of the leading developers of standards in this area, helping to put the maDMP idea into practice through a Research Data Alliance (RDA) working group for Common DMP Standards. That RDA working group recently produced an application profile for maDMP, which has gained approval as an official RDA Recommendation (Miksa, Walk & Neish, 2020).

With the standard approved, and take-up by DMP tool providers increasing, TU Wien have been quick to explore the opportunities to benefit locally. Miksa is now a Senior Scientist with the Center for Research Data Management[1] at TU Wien, which was established to bring together people from different units within the institution, such as the library and the Research Support Office. The Center now employs its own personnel to offer support to researchers, including implementing their DMPs across the whole research data lifecycle.

Coordinating service provision within and across universities

The TU Wien project featured in this story implements the maDMP standard, using a tool called DAMAP, and connects this to other services. The aim is to improve university internal workflows, e.g. allowing reporting on storage needs that have been highlighted in DMPs and using this information to provision for the expected storage needs. As Tomasz explains, "We wanted to see the benefits for those taking care of services, for example reporting on to the management of these services about the provisioning of storage. By connecting systems we can be in a better position to make workflow integration possible, so we can send them DMPs that identify the expected storage needs of each research project."

> *"The gamechanger for us has been the FAIR Data Austria project. Our university and others in Austria had a chance to hire people to implement stuff. This ramped up the speed of the TU Wien project."*
>
> *Dr Tomasz Miksa*

Coordinating service provision within and across universities

The resourcing for the project has come from the national level project FAIR Data Austria . Led by TU Graz, this project aims to support the sustainable implementation of the European Open Science Cloud (EOSC), and is designed to strengthen knowledge transfer between universities, industry, and society. The assumption, says Miksa, is that maDMPs can only be implemented if DMP tools are closely connected with others. This means every university must have a customized and tailored deployment of the maDMP tool, working at different speeds.

DAMAP follows the DMP template guidelines issued by Science Europe, taking their guidance and evaluation criteria into account when supporting the researcher in answering the questions. Answers provided in the tool are stored in a machine-actionable way and the resulting DMP can be exported as a Word document that researchers can further customise and edit before submitting to their funder.
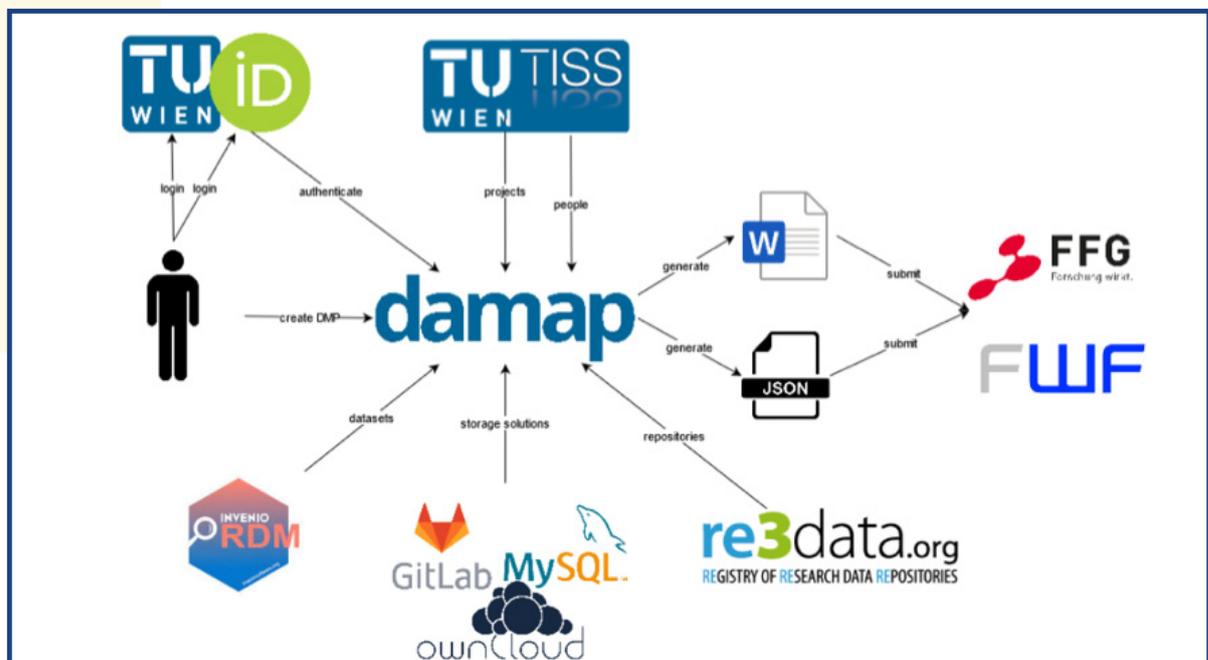
Overview of the DAMAP tool integration with other systems, from Casellato (2021)

With the additional staff in place, the project team tackles implementations between university systems. It will also set up an institutional data repository and launch training. The project team started working with the various teams in the university providing data storage. It has worked with them on details such as the frequency of back-ups and other information that a researcher would need to complete their project DMP, but would not necessarily know about. Collating this information on storage provision has helped the relevant units in the university reflect on their own workflows and market their offers better to researchers. When a researcher now picks one of the storage options in the DAMAP tool, details on data storage are automatically populated for them. This should support researchers in making the right storage choices, to ensure the data can be made FAIR.

The DAMAP tool also aims to use the information in DMPs to inform the deposit of FAIR data via the TU Wien data repository. Designing DAMAP to integrate with repository deposition and ingest workflow involves extracting information provided in the DMP about specific datasets, e.g. on the licensing terms applied, data formats, and volumes, and using this to prefill metadata about the dataset submitted.

▪ Making the most of standards

The tool applies the RDA maDMP application profile, which maps terms associated with data management plans to terms that have been defined in other, more generic, metadata standards. These standards include, for example:

- ORCID[3] for information about the individuals identified in the plan.
- DataCite[4] for metadata associated with the DOI of a dataset identified in a DMP and published in a repository, and to identify its contributors.
- CERIF[5] for information about a research project, to facilitate integration with CRIS systems.

The vision is to use well-defined controlled vocabularies to define the properties of the metadata items that will be exchanged between systems, in some cases these have yet to be standardised.

Stepping up to Integration

▪ Integrating the DMP tool with institutional workflows

Integrating the DMP tool with institutional workflows means dealing with a range of challenges that can be technical, social or organisational. These include:

- Improving semantic interoperability, by adopting controlled vocabularies for the information exchanged between systems, based on open standards.
- Managing expectations as researchers' awareness of available services is raised through integration with DAMAP.
- Bridging the cultural gap between research and administration, through mutual appreciation of what it takes to make DMPs more useful to research activity.

DAMAP's adoption of standards is a work in progress. For example the project team also plans to integrate with more persistent identifier systems, such as ROR[6] for institutions. Challenges arise when other university systems do not use a common standard that can easily be mapped to facilitate the interoperability needed for integration. The university's CRIS for example is a bespoke system developed in-house, so integration needs to be done using its API.

The usability of DAMAP for researchers is a core challenge for the project. There is therefore an overriding need for the project to manage expectations, both on the researchers side and the service providers. Making this tool easily usable for researchers means striking the right balance, at the right time, when asking researchers to identify all the information that funders would like to see in a data management plan.

3. https://info.orcid.org/documentation/integration-guide/orcid-record/
4. DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/3w3z-sa82
5. https://www.eurocris.org/services/main-features-cerif
6. https://ror.org/

Stepping up to Integration

> *"Our goal for machine-actionable DMPs is to make it as easy as possible to get a DMP that the funder will accept, not to educate them about RDM."*

This means keeping the required information limited, structured and standardised, both for usability purposes and in order to be useful for other institutional workflows that include research data.

The timing of these workflows in the research project lifecycle is important. For example, when projects undergo legal and ethical review this usually requires more detailed information from the researcher than the DMP does. But it also needs to be submitted at a later stage than the initial pre-funding DMP that many funders require in research bids. Integration of that information with the DMP might reduce the researchers' effort needed to re-key information, but it would also overload the users' initial interaction with the DMP tool. As Tomasz reflects, "if we incorporate huge questionnaires into the DMP tool, then 'we' will be hated for causing more problems. If we keep the other systems in parallel, then we make people use different tools for everything. This is a trade-off and sometimes the final decision not only must take into account users' experience, but also internal politics".

Conversations are ongoing to identify the right times for service providers to ask the right questions to researchers. The team has learned through its prototyping of DAMAP that there is a demand from researchers to fulfil their needs that may require further mapping between systems, so that information from the DMP can be pushed to other systems that can offer a useful response. The DAMAP team is very conscious that the requirements expressed by researchers in a DMP reflect the collective knowledge of the research group, and DAMAP needs to support multiple users at this level.

The project's discussions of requirements benefit from Miksa's prior experience of research culture. "As a researcher, I see how differences in terminology happen, and this project has helped appreciate these differences." Although it is difficult to say that the meetings held so far contributed directly to better understanding of researchers' needs, "we did our best to break the division into the two worlds of research and administration." The ambition is to uncover more than just 'perceived needs', as prototyping DAMAP will clarify how it may respond to more specific behaviours, so the team can have "discussion on specific workflows, needs, etc. identifying how researchers behave".

The DMP workflow integration with storage provision has led to useful synergies being identified, which may help bring researchers and providers in closer contact. The Centre has taken stock of a wide range of IT services available across the institution. "For DAMAP we needed to have a list of these, and we wanted to have quality of service questions. We asked the providers and they had to figure these questions out for the first time. So this was a nice example of how we get synergies."

These successes bring further challenges in managing demand for services. "We are channelling people to use the services provided by our university," says Tomasz, pointing out that "the services they offer are free at point of use, but if everyone starts using them there are questions about who pays for them."

This challenge of managing expectations and demand for services may be key to the broader question still to be addressed, which is the future funding for this integration, and for the maintenance of the DMP tool and repository systems created.

Impacts

The project is in early stages and additional integrations between systems are to be added, so the full impact of integrating the RDM support workflows at TU Wien is yet to be seen. It is too early to say how researchers' views of the obligation to produce DMPs are affected by the DAMAP tool's ability to respond to these DMPs. Thomas Miksa points out that "people are still curious, as it is a relatively new thing in Austria to require DMPs."

There have, however, already been noticeable changes in awareness of the RDM services available to the researchers. Implementation has increased visibility of the Center for Research Data Management, and highlighted synergies to be leveraged. Reflecting on the lessons for other institutions, Miksa says "implementing maDMPs will make you rethink how you run your current systems and processes. For example, whether information from your CRIS system (or equivalent) can be easily obtained by the maDMP tool.

Sometimes, this can be solved by software developers relatively quickly and there is no need to buy a new system. In some other cases, it can require more work to make a decision on which controlled vocabularies can be used by both systems, or how they could be mapped. All these challenges are not directly maDMP-tool challenges, but they must be solved to improve the RDM processes in which the tool is involved." As the project progresses, further lessons and recommendations will be shared via FAIR Data Austria.[2]

Further information

Blumesberger, S., et al (2021). FAIR Data Austria – Aligning the Implementation of FAIR Tools and Services. Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare. 74, 2 (Nov. 2021). DOI: https://doi.org/10.31263/voebm.v74i2.6379.

Casellato, Z. (2021). DAMAP tool. Presentation given at the RDA VP17 on 22 April 2021. Slides are available at https://www.tuwien.at/fileadmin/Assets/forschung/Zentrum_Forschungsdatenmanagement/pdf-Sammlung/RDA_Zeno_DMAP_20210422.pdf

Miksa, Tomasz, Walk, Paul, & Neish, Peter. (2020). RDA DMP Common Standard for Machine-actionable Data Management Plans. https://doi.org/10.15497/rda00039

Tomasz Miksa, Simon Oblasser, and Andreas Rauber. 2021. Automating Research Data Management Using Machine-Actionable Data Management Plans. ACM Trans. Manage. Inf. Syst. 13, 2, Article 18 (June 2022), 22 pages. https://doi.org/10.1145/3490396

About FAIRsFAIR Implementation Stories

FAIRsFAIR - Fostering FAIR Data Practices in Europe

Acknowledgements

Disclaimer

Copyright rests with the authors