Automating Research Data Management Using Machine-Actionable Data Management Plans

TOMASZ MIKSA, TU Wien & SBA Research, Austria

SIMON OBLASSER and ANDREAS RAUBER, TU Wien, Austria

Many research funders mandate researchers to create and maintain data management plans (DMPs) for research projects that describe how research data is managed to ensure its reusability. A DMP, being a static textual document, is difficult to act upon and can quickly become obsolete and impractical to maintain. A new generation of machine-actionable DMPs (maDMPs) was therefore proposed by the Research Data Alliance to enable automated integration of information and updates. maDMPs open up a variety of use cases enabling interoperability of research systems and automation of data management tasks.

In this article, we describe a system for machine-actionable data management planning in an institutional context. We identify common use cases within research that can be automated to benefit from machine-actionability of DMPs. We propose a reference architecture of an maDMP support system that can be embedded into an institutional research data management infrastructure. The system semi-automates creation and maintenance of DMPs, and thus eases the burden for the stakeholders responsible for various DMP elements. We evaluate the proposed system in a case study conducted at the largest technical university in Austria and quantify to what extent the DMP templates provided by the European Commission and a national funding body can be pre-filled. The proof-of-concept implementation shows that maDMP workflows can be semi-automated, thus workload on involved parties can be reduced and quality of information increased. The results are especially relevant to decision makers and infrastructure operators who want to design information systems in a systematic way that can utilize the full potential of maDMPs.

CCS Concepts: • **Applied computing** → **Enterprise data management; Business process management; IT architectures**; • **Information systems** → **Digital libraries and archives**; • **Social and professional topics** → **Automation**;

Additional Key Words and Phrases: Data management plan, machine-actionable, business processes, enterprise architecture, funder template, requirements engineering, automation, RDM, RDA, FAIR

ACM Reference format:

Tomasz Miksa, Simon Oblasser, and Andreas Rauber. 2021. Automating Research Data Management Using Machine-Actionable Data Management Plans. *ACM Trans. Manage. Inform. Syst.* 13, 2, Article 18 (December 2021), 22 pages.

<https://doi.org/10.1145/3490396>

This research was carried out in the context of the Austrian COMET K1 program and publicly funded by the Austrian Research Promotion Agency (FFG) and the Vienna Business Agency (WAW). This work was also supported by the FAIR Data Austria project funded by the Austrian Federal Ministry of Education, Science and Research.

Authors' addresses: T. Miksa, TU Wien & SBA Research, Vienna, Austria; email: tmiksa@sba-research.org; S. Oblasser and A. Rauber, TU Wien, Vienna, Austria; emails: simon.oblasser@student.tuwien.ac.at, rauber@ifs.tuwien.ac.at.

This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2021 Copyright held by the owner/author(s).

2158-656X/2021/12-ART18

<https://doi.org/10.1145/3490396>

1 INTRODUCTION

The data revolution continues to transform every sector of science, industry, and government [1]. The economic and societal benefits and increased effectiveness of research funding, by ensuring that data generated and (pre-) processed as part of research remains available for re-use, prompted funding bodies and institutions to mandate researchers to write and maintain documents describing responsibilities for data processing and management, known as **data management plans (DMPs)** [46]. A **Data Management Plan (DMP)** is a document to demonstrate awareness of good data practice and is expected to outline how research data is created, managed, shared, and preserved [18]. A **DMP** typically describes, among other things, what types and volumes of research data will be created, how data will be organized and documented, how data quality will be assured, which kind of metadata will be created, how the data will be stored and backed up during the project, how data will be shared, legal and ethical aspects, how data will be curated and preserved beyond the funding, and what budget and resources are needed for implementing the plan [10, 23].

DMPs to date are mostly free-form text documents that are created with text processors or online tools such as DMPonline,¹ and DMPTool² [44]. The tools offer funder-specific DMP templates (e.g., the National Science Foundation [31] or European Commission Horizon 2020 [8]) that mostly present open questions to the researcher and provide guidance depending on the chosen template and the research organization. Some of the tools also provide closed questions where the researcher can choose from a list of controlled fields, such as DS-Wizard³ [17]. Examples of DMPs are available in collections maintained by the Association of European Research Libraries⁴ and openAIRE.⁵

Researchers often perceive **DMPs** as an administrative burden with limited use to write a static document often before the start of a project rather than being an integral part of **research data management (RDM)** [12]. Writing a DMP can be a tedious exercise since there are no means of automation such as pre-filling a DMP with information from related information systems, an automated file analysis for the description of data, automated support for license selection, storage booking, data deposit, or cost estimation for data management resources [28]. Recipients of a DMP such as funders do not have automated support either and have to manually review them [42]. Other stakeholders involved in the process of RDM, such as repository operators, are typically not involved in data management planning from the beginning, making it difficult for them to assist with data curation and enable hosting/provisioning of data via their repositories during and after research projects. ICT operators who provide computing resources and storage are not well integrated into data management planning and are often not informed about upcoming service demands to plan their resources efficiently [42]. Institutions are increasingly offering training in data management and the according planning steps to their researchers, and are establishing data stewardship positions to support them. However, to date, these activities mostly do not result in solid data management planning processes, integrating information from various stakeholders, and reducing DMP activities to a negligible burden for researchers while still resulting in high-quality, effective DMPs. Another major disadvantage arises from the current format of a DMP that is mostly a text document not favorable for machine processing and therefore hinders to exploit the information for data management activities, such as monitoring, reporting, or validation.

¹<https://dmponline.dcc.ac.uk>.

²<https://dmptool.org>.

³<https://ds-wizard.org>.

⁴<https://libereurope.eu/working-group/research-data-management/plans/>.

⁵<https://hdl.handle.net/11353/10.1140797>.

To sum up, a DMP in its current form does not reach its full potential as a satisfactory, integral part of research, which enables the flow of information and the automation of data management processes from which many stakeholders can benefit.

The **Research Data Alliance (RDA)**, which is an international organization promoting data-sharing and data-driven research, identified the limitations of current DMPs and took steps toward “active” DMPs [43]. A standardized machine-actionable (meta-)data model [14, 19, 39] for DMPs is a prerequisite for DMPs to become “active” and to enable the exchange of DMP information between research systems and the automation of RDM tasks throughout the research data lifecycle. The RDA DMP Common Standards working group produced an official recommendation [29] that describes an application profile for **machine-actionable DMPs (maDMPs)**. Thus, the RDA established a common way to model information that is typically described in DMPs.

However, to fully realize the potential of maDMPs as a way to exchange and act on information about data used and produced by researchers, all stakeholders involved in RDM must be connected by information systems that manage and exchange information in an automated way. For that purpose, services that provide structured machine-actionable information, which can be fed into a DMP, and services that can consume this information and act upon it are required.

In this article, we design a domain-agnostic system that utilizes the RDA’s recommendation on maDMPs. We explore the feasibility of automation of typical data management processes using maDMPs in an institutional context, such as a research institution or university with its systems, services, and stakeholders. This involves (i) identification of relevant stakeholders at a research institution and related organizations (e.g., funders) and their requirements for a machine-actionable data management planning support system, (ii) development and description of workflows/business processes for machine-actionable data management planning, (iii) description of the system architecture on an enterprise level that can serve as a reference architecture for maDMPs at institutions, and (iv) development of a proof-of-concept implementation that shows selected features of the described architecture.

We evaluate the proposed system in a case study at TU Wien, which is the largest technical university in Austria, by implementing selected workflows and associated services of the architecture as a proof-of-concept, connecting with the institutional infrastructure of systems and services of TU Wien. We follow the design science methodology [13, 33] and use the proof-of-concept implementation to evaluate the DMP efficiency gain by assessing the degree of automation achieved for data management planning tasks. We also assess the effectiveness with which the resulting maDMP can meet the stakeholder requirements. To evaluate how the maDMP meets funder requirements, we use the DMP templates of the European Commission and a national funding body in Austria and assess to which extent they can be covered.

The information system described in this article shows how machine-actionable data management planning can be embedded into the landscape of an institutional research data management infrastructure. Thus, our work helps decision makers and infrastructure operators at research institutions design in a systematic way information systems that can utilize the full potential of maDMPs.

The article is structured as follows. Section 2 describes related work. Section 3 describes the requirements engineering carried out for this work. Section 4 describes the proposed architecture for an institutional maDMP support system. Section 5 describes the proof-of-concept implementation of selected workflows and services of the proposed architecture in a real-world environment. Section 6 evaluates the degree of automation achieved for DMP tasks with the proof-of-concept implementation. Section 7 provides the conclusion and future work.

2 RELATED WORK

maDMPs are a new generation of DMPs. The central characteristic of an maDMP is that its information is modeled in a machine-actionable way. The term *machine-actionable* is associated with the **Findable, Accessible, Interoperable, Reusable (FAIR)** principles to express that machines should be able to autonomously take action on digital objects [48]. Machine-actionability for DMPs shall be achieved by modeling the semantic information with the use of controlled vocabularies and standards [26], as well as by using **persistent identifiers (PIDs)** to reference specific entities, such as people, institutions, funders, grants, datasets, or repositories [41, 42].

RDM still has low maturity at many research institutions [9]. Universities maintain data in several information silos, each of them engineered to serve a specific vertical application. Data about key entities, such as people, publications, courses, and projects, is scattered across them and difficult to correlate due to the diversity in format, metadata, conventions, and terminology used [22]. When a DMP is machine-actionable, it can become a living (active) document [28] since information can be integrated from various sources [27] and automatically be updated. maDMPs facilitate the exchange of information across research systems and enable new use cases as described in the work of Simms et al. [42]. Transition from DMPs to maDMPs can be seen as a transition from *Level 1 Content Management* to *Level 3 Knowledge Management* as defined in the work of Tuzhilin [47]. This is because maDMPs model the actual information and are not simple text documents that structure is hard to interpret by machines. The benefits of using structured information over traditional text-based approaches to automate document management was discussed in the work of Lee et al. [21].

The RDA DMP Common Standards working group [39] developed a common data model/metadata application profile for DMPs [29]. The recommendation was developed as a collaborative effort by consulting stakeholders and collecting user stories as described in the work of Miksa et al. [26]. It is capable of describing various entities involved in data management planning such as the DMP itself, projects, funding, contributors, costs, datasets, and their relations. An overview of the RDA DMP Common Standard data model is given in Figure 1.

The principles described in the work of Miksa et al. [28] specify what is required to bring the maDMP vision to life and to realize use cases from stakeholders involved in RDM. Among other things, the principles suggest embedding maDMPs in the workflows of stakeholders to enable automation and to use a common data model for the exchange of DMP information. Further automation can be enabled by the machine-actionable description of resources such as data policies or components of the research data ecosystem such as repository systems. maDMPs do not replace other systems for managing data [30, 38], storing data [3], or citing data [16] but complement them to establish research data services [20, 45].

The research community also identified potential use cases [2, 41, 42] and requirements [26] for maDMPs and supporting systems. This article breaks down these use cases into specific processes and describes how they can be realized in an institutional context.

3 REQUIREMENTS ENGINEERING

Requirements engineering plays a crucial role in the development process of an information system as it aims at providing a complete and accurate requirement specification [6]. In this section, we describe how we collected requirements for the information system.

Since there were no well-established use cases for **machine-actionable Data Management Plans (maDMPs)**, we started our work in 2017 by identifying specific requirements. We co-organized workshops and consultations to identify together with the scientific community typical use cases for **maDMPs**. Specifically, the outputs from the IDCC workshop held in Edinburgh in

Fig. 1. Overview of concepts used in the RDA recommendation on maDMPs [29].

2017 that gathered almost 50 participants from Africa, the United States, Australia, and Europe form the basis of our work [42]. We have further refined the requirements in physical workshops and virtual meetings. Together with the RDA DMP Common Standards working group, we co-organized between October 9, 2017, and November 30, 2017, an open consultation in which 108 user stories were collected. They express viewpoints of funders, institutions, repository operators, research support, researchers, and service providers [26]. The participants contributing their user stories represented different institutions, such as universities, medical universities, universities of economics, technical universities, and social data archives [24]. Thus, the requirements identified in the consultations are neither limited to the use case discussed in this work nor to particular domain, such as a technical university.

Based on the requirements, we developed workflows (Section 3.1) and graphical mockups (Section 3.2) that reflect these use cases. The workflows underwent an initial evaluation during a workshop hosted during the International Conference on Theory and Practice of Digital Libraries in 2018 [25]. A group of about 15 participants was divided into breakout groups and asked to review a series of workflows that could potentially be automated under an maDMP ecosystem. Each group had to answer the following questions:

- What would you change in the workflows?
- Do the workflows fit into the context of your organization/domain?
- Which other stakeholder needs should be addressed?
- Which other systems can be used?
- What else could be automated?

The mockups were tested and refined with the help of relevant stakeholders at TU Wien but also external bodies such as a representative of a major research funding agency in Austria.

We are aware that more requirements engineering needs to be conducted. However, with our exploratory work, we hope to advance the discussion and contribute to a better understanding of **maDMPs** and the surrounding processes.

3.1 Automated Workflows

The design of the workflows is framed by the process of creating an initial DMP at an early stage of a project [26] but also considers data management tasks at later stages, such as when data is ready for submission into a repository system [28].

Fig. 2. (a) Overview on a high-level workflow of creating a DMP at an initial phase of the project. (b) Stakeholder services use the **maDMP** as a medium for information exchange.

Figure 2(a) depicts a high-level workflow of creating an initial DMP and shows that services and stakeholders of an RDM infrastructure are involved in the process [26]. In this high-level workflow, the researcher starts a DMP and administrative information like researcher affiliation is fetched from related information systems to pre-fill the DMP. Next, the researcher estimates the expected size and type of the research data and gets assistance in storage booking, cost estimation, and license selection for datasets. The DMP is a living document that evolves over time. Hence, especially in the initial DMP, the information cannot be considered final and its correctness may be subject to review of other stakeholders—for example, estimations of required storage made by researchers may need to be revised by the ICT operator.

Figure 2(b) shows potential interactions between stakeholder services using an **maDMP** as a medium for information exchange at a later stage in the project [28]. Information contained in the **maDMP** like the specified license or embargo period for a dataset can be used to facilitate the submission process of research data to the selected repository by automatically assigning these fields in the repository system. A repository operator service can return a list of **Persistent Identifiers (PIIDs)** pointing to the submitted datasets and provide associated cost to update the **maDMP**. A funder service can use the information to check how the DMP was implemented.

The high-level workflows described so far remained vague, and it was not clear what exactly should be done at each step. Hence, we modeled nine sub-workflows using the **Business Process Model and Notation (BPMN)**. Figure 3 gives an overview on the high-level **Business Process Model and Notation (BPMN)** workflows, which form the foundation for our system architecture design. We describe all workflows in a technical report [36]. In the following, we present one of the workflows.

The *Specify Size and Type* workflow, illustrated in Figure 4, deals with the specification of research data that will be (re-)used and generated during the project:

- (1) *Cite data*: If existing, citable datasets are being reused, their unique and resolvable **PIIDs** like DataCite⁶ **digital object identifiers (DOIs)** can be entered and associated metadata be retrieved and imported into the DMP. Many repositories support the **Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)** standard, and other repositories like

⁶<https://datacite.org/doi.html>.

Fig. 3. Overview of the automated workflows using BPMN.

Fig. 4. Specify size and type of research data.

GitHub provide a REST-API to collect metadata. Citing reused datasets in a DMP enables the data discovery use case for datasets referenced in different DMPs. Retrieved metadata such as the license of the reused dataset clarify the terms of use [5] and can avoid pitfalls already in the planning phase.

- (2) *Specify output data*: The researcher can specify which data will be generated during the project. The researcher can specify the expected types, file formats, and volumes of research data manually or get automated support by uploading datasets that are analyzed (file format, size, other metadata) by a file identification and characterization tool. If entire datasets are already known or available, they can be uploaded for analysis. Alternatively, sample data can be uploaded and analyzed, the expected number of similar files can be specified, and an estimate of the total size can be calculated.
- (3) *Classify/tag data*: The researcher can classify/tag the data, such as by assigning labels.
- (4) *Save*: The workflow concludes with storing the details of the research data specification to the DMP.

3.2 Graphical Mockups

In Section 3.1, we described automated workflows for data management planning using BPMN and textual descriptions. BPMN is a notation used to facilitate discussions about business

Machine-actionable Data Management Planning Application

https://dmap.tuwien.ac.at

DMap | MyDMPs

Notifications | John Doe | Sign Out

Specify your research data | Get Help

What kinds of research data will you create?

☐ Don't know yet

☐ No data will be created or analysed

☒ Specify with assistance

Create tags to logically group your research data into datasets.

Supplementary material Source code for client application Raw data Processed data Add tag...

Estimate type, volume and format of your research data manually

Estimated type Estimated size Dataset tag

Text documents (doc, odg, pdf, txt, etc.) < 100 MB Default Add

or by uploading sample data, similar to the data you will create.

Choose files to upload or drag and drop them here

Estimation

Filename	Estimated type	Size	Dataset tag	Nr of similar files	Estimated total size
-	Software application source code	100 - 1000 MB	Source code for client application	-	100 - 1000 MB
pic1.tiff	Graphics/Images	4.3 MB	Raw data	1000	1 - 5 GB
track1.wav	Audio	10.5 MB	Raw data	300	1 - 5 GB
result.csv	Spreadsheets	915 kB	Processed data	150	100 - 1000 MB
script.sh	Text documents	234 kB	Default	10	< 100 MB
report.pdf	Text documents	5.9 MB	Supplementary material	5	< 100 MB
screen.png	Graphics/Images	2.3 MB	Default	100	100 - 1000 MB

Fig. 5. Wireframe for specifying research data.

processes between non-technical and technical people by using a common language. However, from a **user interface (UI)** design perspective, it is not straightforward to derive requirements for a **User Interface (UI)** from the **BPMNs**. UI mockups present a quick opportunity to test concepts with stakeholders and receive valuable feedback on the further design. Therefore, we developed interactive, graphical UI mockups [35] using Balsamiq⁷ for five different stakeholder groups:

- (1) Researcher
- (2) Research Support
- (3) **Information and Communications Technology (ICT)** Operator
- (4) Management
- (5) Funder

All of these stakeholders are involved in the data management planning workflows and related use cases. The graphical UI mockups are composed of about 60 wireframes. The wireframes are linked with each other and contain interactive, clickable elements. We published the graphical mockups on GitHub pages⁸ as a PDF and with a link to the interactive version hosted at Balsamiq Cloud. Visitors of Balsamiq Cloud could provide comments with markers directly on the wireframes. Figure 5 shows a sample wireframe designed for a researcher. It allows researchers

⁷<https://balsamiq.com/>.

⁸<https://doi.org/10.5281/zenodo.3630374>.

to specify what data will be used by defining tags that represent a dataset, such as “Raw data.” Researchers describe the dataset by selecting from pre-defined options, such as information on type, size, and format. Researchers can also simply upload sample files instead. The system will detect file format and size. The user will have to specify how many “similar” files will be used/produced in the project.

Additionally, we met in person with representatives from the stakeholder groups located in Austria and presented the mockups and documented their feedback.

The feedback collected contributed to the improvement of the graphical mockups. In this work, we went through several iterations of incorporating feedback, with v1.2 [35] being the latest version of the mockups.

4 ENTERPRISE ARCHITECTURE

In this section, we introduce an architecture for machine-actionable data management planning in the context of a research institution and its infrastructure. A research institution or university with its various stakeholders, systems, and services can be viewed as an enterprise. We therefore use a software engineering technique called **enterprise architecture (EA)** modeling to describe the architecture from various perspectives, highlighting different aspects. The purpose of the architecture description is to give a comprehensive overview of the system by using different architectural views to highlight different system aspects, capture significant architectural decisions, and communicate them to various stakeholders.

In our work, we defined architectural goals and associated requirements and constraints by assessing the drivers for change of an institutional **Research Data Management (RDM)** infrastructure. We integrated the workflows derived from community use cases into the business process layer of the architecture. For the business processes and sub-processes, we derived application services and components that can implement them. By arranging architectural components around business capabilities, we achieved a modular architecture that is polyglot and enables self-contained components to be independent of programming language and technology.

Due to space limitations, we cannot present all the views. The full description of the architecture with all its views (use case view, logical view, process view, implementation view, deployment view) can be found in the work of Oblasser [34]. In the rest of this section, we describe the main building blocks of the proposed architecture.

Figure 6 shows the business process views of our architecture. In Figure 6(a), the stakeholders and the business services are depicted—for example, the *Data Management Planning* service serves the researcher and is realized by the *Data Management Planning* business process. The *Data Management Planning* business process consists of sub-processes, which are listed in Section 3.1.

The *Get Storage* business process implements the storage provisioning for the researcher. The *Get Help* business process realizes the *DMP Support service* for the researcher and research support.

The result of the *Data Management Planning* business process is the *DMP* business object, representing the actual DMP. Data management cost is an output of the *Configure Storage and estimate Cost* sub-process.

Figure 6(b) shows business services of the ICT operator and the library. The *Storage Demand Overview service* serves the ICT operator. The library is served by the following business services: *Preserve Data service*, *Open Access Publishing service*, and *Issue DOI service*.

The layered view depicted in Figure 7 shows the relations between the elements of the business layer (yellow), the application layer (blue), and the technology layer (green). Each layer provides services that serve the layer above. In this way, it can be traced which business process is served by which application service and implemented by which application component. Further, it can

(b) ICT operator and library served by their business process services.

Fig. 6. Business process views of the proposed architecture depicted using Archimate notation.

Fig. 7. The layered view (ArchiMate) of the proposed architecture: business (yellow), application (blue), technology (green).

Table 1. Proposed Web Application Services for this **Enterprise Architecture (EA)**

Proposed Service	Description
DMP Store	Web service maintaining a repository for DMPs providing an API for operations like searching, creating, accessing, modifying, and deleting DMPs
DMP App	Web application serving the DMP GUI and implementing DMP logic
Repository Recommender	Web service for recommending repositories (API)
Metadata Standard Discovery	Web service for discovering metadata standards (API)
Metadata Importer	Web service for importing metadata (API)
File Characterizer	Web service for uploading and analyzing file samples, identifying their file formats, and providing characterization metadata (API)
Notifier	Notification service to deliver messages to users (API)
Administrative Data Collector	Web service for importing administrative data from information systems such as CRIS or ORCID (API)
Repository Ingestor	Web service for ingesting data into a supported repository (API)
Service Broker	Web service for brokering services of an ICT service provider; the broker provides a catalog of services and enables their provisioning (API)
Service Catalog Controller	Web service for registering service brokers from different providers and providing an aggregate catalog of available ICT services (API)
Help Desk	Web application serving the help desk GUI
ICT Dashboard	Web application serving the ICT dashboard GUI

be traced which application components are served by which technology service and realized by which technology.

The illustrated application layer only partly shows how application components interact. For example, the *CRIS API* serves the *Administrative Data Collector* component. In other words, the *Administrative Collector* component fetches administrative data from the *CRIS API*. Details about how application components are integrated and how data flows between application components are depicted in the implementation view described in the work of Oblasser [34].

The technology layer (green) at the bottom of Figure 7 suggests that application components are realized as web application services. Some web application services are intended to be stateless such as the *Metadata Importer*, whereas others require a database to persist data such as the *DMP Store*. We identified 13 application services that are required to implement the business processes (Table 1). The services vary in complexity and form. Some services are intended to provide a GUI (graphical user interface), whereas others offer an API (application programming interface). The system is designed to be deployed on an institutional level, but some of the services could be shared with other institutions or the public. For example, the *Repository Recommender* service that consumes DMP and policy information in a standard format can be deployed publicly. The same applies for the *Metadata Standard Discovery*, the *File Characterizer*, the *Metadata Importer*, or the *Repository Ingestor* services under certain conditions. The *Help Desk* service could be shared between institutions.

5 IMPLEMENTATION

In this section, we describe an implementation of the proposed maDMP support system in an institutional context. We chose TU Wien as an example institution to show how the proposed system can be embedded in the system and service landscape of an institution.

This section is structured as follows. First, we introduce the TU Wien case study. Second, we present the DMap tool, which implements selected automated workflows using the TU Wien infrastructure and other systems and services.

5.1 TU Wien Case Study

TU Wien is a major university in Austria with more than 27,000 students, around 5,000 employees, and eight faculties. It has departments, information systems, and **ICT** services that are relevant to our maDMP use cases.

The automated workflows described in Section 3.1 involve three main stakeholders, namely the researcher, research support, and **ICT** operator. At TU Wien, DMP-related research support is offered by the Center for Research Data Management and the European and International Research Support (EIRS) team. TU.it is the in-house **ICT** operator that provides **ICT** services for employees and researchers. The services offered range from storage solutions to virtual server hosting to the provision of **High Performance Computing (HPC)** infrastructure and more.

The **TU Wien Information Systems and Services (TISS)**⁹ is the central information system to ensure university operations. TU Wien Information Systems and Services (TISS) includes a searchable address book of employees, students, and organizational units. TISS also contains a project database to manage research projects and their funding. Both of these two sub-systems contain relevant information for DMPs. The address book serves as a source for personal information about researchers and other employees, such as person ID, email, or affiliation. The project database contains relevant project and funding information, such as project ID, title, duration, funder, or grant ID. TISS provides a public **REpresentational State Transfer (REST)** API to fetch information from the address book and the project database.

5.2 DMap Tool

To show the applicability of the EA described in Section 4 in a real scenario, we developed a proof-of-concept tool named *DMap*.¹⁰ DMap uses the infrastructure of TU Wien and its systems and services. In DMap, we implemented selected use cases from the set of workflows described in Section 3.1. DMap is the implementation of the *DMP App* application service and related application components described in the EA and therefore realizes the service used by researchers and research support staff. Other application services described in the EA such as the *ICT Dashboard* or the *Help Desk* are not implemented as part of this work. The DMap UI was implemented according to the mockup design described in Section 3.2. Figures 8 and 9 show screenshots of DMap in action. Further, DMap screenshots can be viewed in the slides [33] resulting from the presentation of DMap to the Open Science community at the 15th RDA plenary in Helsinki in October 2019.

DMap is separated into two decoupled services - a backend and a frontend service. The communication between the frontend and backend is realized as a resource-based, synchronous communication over HTTP (Hypertext Transfer Protocol). Hence, the backend provides a **REpresentational State Transfer (REST) Application Programming Interface (API)** that serves as a gateway to various kinds of resources, such as DMPs stored in the database or project information located at the external TISS. JSON (JavaScript Object Notation) is used as the payload format for the

⁹<https://www.tiss.tuwien.ac.at/>.

¹⁰<https://github.com/oblassadors/dmap>.

Fig. 8. Project selection in DMap.

Fig. 9. Steps to create a DMP in DMap.

transfer of structured information between the backend and the frontend. The frontend runs in a web browser and provides the UI to the user.

The backend is implemented as a Spring Boot¹¹ application with an embedded Apache Tomcat servlet engine using Java 8. The communication with the DMP database is done via an abstraction data access layer provided by Spring Data. Currently, DMap uses the document database MongoDB¹² to persist DMPs, but it can easily be replaced with any other datastore due to the used data access abstraction layer.

The frontend has been developed with the JavaScript framework Vue.js¹³ and makes use of several libraries to provide a scalable and maintainable architecture. For example, Vue Router is used for client-side routing between different views, which enables us to create a modern single-page web application. The UI is composed of self-contained and reusable Vue components that contain HTML (Hypertext Markup Language) template code, JavaScript code, and Cascading Style Sheets (CSS)-style information. A central Vuex state store is used to maintain a consistent state of the frontend application. To communicate with the backend, a Hypertext Transfer Protocol (HTTP) service using the Axios library has been implemented.

The RDA DMP Common Standard for maDMPs is the native format for the tool. Thus, the *information* is modeled and structured from the beginning. This is different when compared to other tools that mostly structure *questionnaires* (i.e., consist of question-answer hierarchies).

We implemented workflows to assist with the specification of research data, license, and repository selection. DMap can export a maDMP as **JavaScript Object Notation (JSON)** that complies with the **Research Data Alliance (RDA)** DMP Common Standard. We implemented a service

¹¹<https://spring.io/projects/spring-boot>.

¹²<https://www.mongodb.com>.

¹³<https://vuejs.org>.

Table 2. Workflows Implemented in DMap

Workflow	Implemented in DMap
Start DMP	Yes
Specify Size and Type	Yes
Get Cost and Storage	Partially
Get License	Yes
Get Metadata Standard	No
Get Repository	Yes
Deposit Data	No
Get Help	No

Yes, implemented; No, not implemented; Partially, implemented but not integrated with the tool.

broker that represents a standard interface to the services provided by TU.it, the local **ICT** operator. We also integrated the CRIS of TU Wien to fetch structured information about projects and researchers. Table 2 provides an overview of implemented business processes (cf. Section 3.1).

We introduced some simplifications of implementation compared to the original specification of workflows and architecture description. Specifically, we did not model machine-actionable data policy elements, described as one of the 10 principles of maDMPs [28]. We also simplified the proposed repository recommendation service by integrating a repository search in DMap, backed by the repository registry **Registry of Research Data Repositories (re3data)**, rather than implementing an actual recommender system. To implement a repository recommendation service as described, relevant maDMP metadata must be aligned with repository metadata and machine-actionable data policies, and a suitable recommendation model must be developed. The simplifications have no effect on the later evaluation.

6 EVALUATION

In this section, we describe how we evaluate the proposed information system for maDMPs. The implemented proof-of-concept tool DMap described in Section 5.2 serves as a vehicle for our assessment to make statements about the usefulness of the system.

In Section 6.1, we assess the system efficiency by evaluating the processes implemented in DMap. For this purpose, we analyze the extent to which the creation of a DMP in DMap is automated and simplified compared to the conventional way of writing a DMP in free form.

In Section 6.2, we evaluate the maDMP created with DMap with regard to its completeness of information that is relevant for funders. The funders' requirements for DMPs are currently expressed in funder-specific DMP templates. Therefore, we evaluate how well the questions of two major DMP templates (the **Austrian Science Fund (FWF)** [4] and the European Commission's Horizon 2020 programme [11]) can be answered with the information contained in their machine-actionable counterpart.

6.1 Degree of Automation

We analyzed the processes involved in creating a DMP from the perspective of a researcher who has to write a DMP for a research project. We set the conventional way of writing a DMP in free-form as the baseline for our analysis. In the analysis, we focused on the key steps of creating a DMP. For each of the steps, we assessed the degree of automation achieved with DMap. The results of the assessment are summarized in Table 3.

Table 3. Steps of Creating a DMP with DMap and Their Assessed Level of Automation

Step	Manual	Semi-Automated	Fully Automated
1.1 Select research project(s)	☐	☒	☐
1.2 Select people involved in data management	☐	☒	☐
1.3 Assign contribution role(s)	☒	☐	☐
2.1 Group datasets	☒	☐	☐
2.2 Estimate research data	☒	☐	☐
2.3 Analyze files	☐	☒	☐
2.4 Calculate total storage size per dataset	☐	☐	☒
3.1 Select datasets to share	☐	☒	☐
3.2 Select license(s)	☐	☒	☐
4.1 Select repository/repositories	☐	☒	☐
4.2 Assign datasets to be deposited	☒	☐	☐

The evaluation shows that most of the processes are semi-automated. Compared to the conventional free-form way of writing a DMP, DMap presents an interactive way of creating a DMP. Instead of being faced with a blank page, the user gets assistance for every step. This is achieved by integrating external information systems (TISS, re3data) and tools (FITS, EUDAT license selector) into DMap, providing a suitable UI (Section 3.2) for various tasks and offering the user pre-defined options to choose from.

By integrating external services and tools, we simplified the user experience by providing a single UI instead of switching between different UIs. A positive side effect is that the data from external information systems is imported into the tool in a controlled manner by interacting with the respective service APIs. This enables to unambiguously reference entities such as persons (TISS ID), projects (project ID), grants (grant ID), licenses (uniform resource identifier), or repositories (re3data ID) in a maDMP. A manual approach to referencing entities is prone to errors, due to possible typing errors.

The Horizon 2020 DMP template survey results [15] show that users wish for more guidance in terms of recommendations and drop-down options to choose from. In our earlier work [37], we discuss a recommender system for repositories. However, in DMap, we did not implement a repository recommender system but provided the user with the ability to find suitable repositories by applying filters and text search. Therefore, there is higher potential for automation than realized in DMap. DMap provides drop-down options for the assignment of contribution roles, the specification of research data types, sizes, and the selection of licenses. We also set default values to simplify the user experience. For example, the licenses for datasets are set to CC-BY by default, or the project end date is automatically set as the date from when the license should be active. However, further drop-down options can be provided, such as a list of suitable metadata standards to select from.

The selection and automated cost estimation for active data storage (*Get Cost/Storage* workflow) was not integrated with DMap and therefore could not be evaluated. However, the implementation of the TU.it service broker (not described in this article) shows that we can implement a standard interface for the interaction with ICT services that enables us to configure ICT services and calculate the costs based on the respective cost models. By integrating service brokers into a data management planning application, we can automate the discovery of suitable ICT services by showing a catalog of services to the user and provide semi-automated assistance in service configuration and cost estimation.

Fig. 10. Analysis results for both DMP templates. More than 80% of the questions of both DMP templates can be answered completely or partially with the information contained in the maDMP.

In this evaluation, we assessed the level of automation from the perspective of a researcher. However, there are other stakeholders of a maDMP support system who can benefit from automated processes. For example, the ICT operator can get automated support in the management of ICT resources by being integrated into the data management planning workflow. The Research Office can be notified if an ethics approval is needed. The funder can semi-automatically evaluate maDMPs because of their machine-actionable format. At the time of writing, these use cases remain aspirational and need to be evaluated after they are implemented.

6.2 Completeness

As funders are the primary recipients of DMPs, we evaluate how well an maDMP can meet the funders' requirements. For this evaluation, we created an maDMP in the RDA DMP Common Standard v1.0 with our proof-of-concept tool DMap. We then used the DMP templates of two major funders (the FWF [4] and the European Commission's Horizon 2020 programme [11]), which contain sets of questions and evaluate to what extent they can be answered with the information contained in the maDMP.

For each of the 31 questions in the H2020 template and 27 questions in the FWF template, we assess whether the question can be answered completely/partially/not with the information contained in the maDMP. If the question can be answered completely or partially, we list the corresponding maDMP fields. Thus, we created manually a table with a mapping between questions from the templates and fields in the maDMPs.

Table 4 presents the results for the FWF DMP template grouped by their categories. Figure 10 depicts the analysis results for both DMP templates. The analysis shows similar results for both DMP templates, although the questions differ in topic and granularity. More than half of the questions (58.1% H2020, 59.3% FWF) of both templates can be completely answered with the information contained in the maDMP. Of these questions, many can be answered with the type-safe, machine-actionable information in the maDMP, such as PIDs, uniform resource identifiers, numerical values, or data fields with a controlled vocabulary, whereas others can be answered with the free-form text fields of the maDMP. For example, the maDMP uses DOIs for all data objects, and ORCIDs and internal staff identifiers for all actors involved.

About one-fifth to one-quarter (22.6% H2020, 22.2% FWF) of questions from both templates can be answered partially, and about one-fifth of the questions (19.4% H2020, 18.5% FWF) cannot be answered. One reason for this is that some of these questions require human narrative for which the maDMP does not provide suitable data fields or the information in the maDMP is not

Table 4. Results of the Evaluation Exercise, Showing the Degree to Which the 27 Questions of the FWF DMP Template [4] Can Be Answered by the Information Contained in the maDMP

FWF DMP Question Category	Question Answerable		
	Completely	Partially	Not
Data Officer	100.0% (1)	0.0% (0)	0.0% (0)
I.1 Description of the data	25.0% (1)	25.0% (1)	50.0% (2)
II. 1 Metadata standards	100.0% (1)	0.0% (0)	0.0% (0)
II.2 Documentation of data	0.0% (0)	33.3% (1)	66.7% (2)
II.3 Data quality control	100.0% (2)	0.0% (0)	0.0% (0)
III.1 Data sharing strategy	100.0% (3)	0.0% (0)	0.0% (0)
III.2 Data storage strategy	71.4% (5)	28.6% (2)	0.0% (0)
IV.1 Legal aspects	25.0% (1)	50.0% (2)	25.0% (1)
IV.2 Ethical aspects	100.0% (2)	0.0% (0)	0.0% (0)
	59.3% (16)	22.2% (6)	18.5% (5)

meaningful enough to adequately answer the questions. For example, the question “How are you planning to document this information?” in the Documentation and Metadata section of the FWF template has no matching field in the maDMP.

If the funder requirements for DMPs do not converge, it is difficult to design a data model that supports all kinds of questions from different DMP templates. The maDMP specification contains a minimal set of common fields used among different templates and use cases. There are developments to harmonize the requirements for DMPs among the funding organizations by establishing core requirements for DMPs [40]. Another solution can also be development of an extension to the maDMP specification, which extends it with funder specific fields that are missing in the core specification. However, this can result in the deluge of extension and lack of interoperability, which in turn contradicts the concept of a common standard for maDMPs.

We performed the evaluation on the templates that are most common for the considered use case. However, it is possible to map maDMPs to other templates. For example, during the maDMP hackathon on maDMPs¹⁴ organized by the RDA DMP Common Standards working group, one of the groups managed to achieve 100% accurate mapping for the National Science Foundation template in the United States [7]. This shows that the maDMPs can be used in non-European context as well.

7 CONCLUSION AND OUTLOOK

maDMPs are a way to exchange information about data used and produced by researchers. Research institutions can build their research data infrastructure around maDMPs to bring researchers together with departments and services. Thus, they can integrate all stakeholders involved in RDM by connecting their information systems to manage and exchange information in an automated way. By automating tasks for data management planning, researchers can be supported on their journey to good data management practice. For that purpose, services that provide structured machine-actionable information, which can be fed into an maDMP, and services that can consume this information and act upon it are required.

In this article, we explored use cases of maDMPs in the context of a research institution or university with its systems and services and proposed an architecture for an information system

¹⁴<https://www.rd-alliance.org/rda-hackathon-madmps>.

that helps in machine-actionable data management planning. We described a system architecture on an enterprise level that can serve as a reference architecture for other institutions that plan to automate data management. A proof-of-concept implementation shows that the use cases are feasible and tasks can be semi-automated, but human interaction and free-form text cannot be evaded completely.

The benefits and limitations of the proposed solution can vary depending on the context of specific deployment, such as the number of implemented processes or services. The effort and expenses depend on the starting point of the specific institution. The EA and processes described in this article can help all stakeholders identify needed changes in the existing services and better coordinate investments to create value for everyone involved in RDA.

maDMPs are still in the early adoption phase, and more research is required. In this work, we focused on providing automated support for the creation of maDMPs and enabling related use cases of institutional stakeholders. Future work will elaborate on the use cases of the research funder, including DMP monitoring and automated support for reviewing DMPs.

REFERENCES

- [1] Serge Abiteboul and Julia Stoyanovich. 2019. Transparency, fairness, data protection, neutrality: Data management challenges in the face of new regulation. *Journal of Data and Information Quality* 11, 3 (June 2019), Article 15, 9 pages. <https://doi.org/10.1145/3310231>
- [2] Data Documentation Initiative ADMP Working Group. n.d. Active Data Management Plans: A Metadata-Driven Model for Data Management Plans (Draft White Paper). Retrieved March 14, 2020 from <https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/7864356/>.
- [3] Ricardo Carvalho Amorim, João Aguiar Castro, João Rocha Da Silva, and Cristina Ribeiro. 2017. A comparison of research data management platforms: Architecture, flexible metadata and interoperability. *Universal Access in the Information Society* 16, 4 (Nov. 2017), 851–862. <https://doi.org/10.1007/s10209-016-0475-y>
- [4] Austrian Science Fund. 2019. FWF Data Management Plan Template (DMP)—Guide (01/2019). Retrieved Mary 14, 2020 from https://www.fwf.ac.at/fileadmin/files/Dokumente/Open_Access/FWF_DMPTemplate_e.pdf.
- [5] Alex Ball. 2014. *How to License Research Data*. DCC How-to Guides. Edinburgh, Scotland: Digital Curation Center. <http://www.dcc.ac.uk/resources/how-guides>.
- [6] Ilyès Boukhari, Stéphane Jean, Idir Ait-Sadoun, and Ladjel Bellatreche. 2018. The role of user requirements in data repository design. *International Journal on Software Tools for Technology Transfer* 20, 1 (Feb. 2018), 19–34. <https://doi.org/10.1007/s10009-016-0443-0>
- [7] Joao Cardoso, Sarah Jones, Tomasz Miksa, Adil Hasan, Maria Praetzelis, Paulette Lieby, Elli Papadopoulou, and Kerstin Gierend. 2020. Mapping of maDMPs to Funder Templates. Retrieved November 9, 2021 from <https://doi.org/10.5281/zenodo.3944458>
- [8] European Commission. 2016. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*. Technical Report Version 3.0. European Commission, Directorate-General for Research & Innovation. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- [9] Andrew M. Cox, Mary Anne Kennan, Liz Lyon, and Stephen Pinfield. 2017. Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology* 68, 9 (Sept. 2017), 2182–2200. <https://doi.org/10.1002/asi.23781>
- [10] DCC. 2013. *Checklist for a Data Management Plan*. Technical Report v.4.0. Edinburgh, Scotland: Digital Curation Center. <http://www.dcc.ac.uk/resources/data-management-plans>.
- [11] European Commission. 2018. *H2020 Templates: Data Management Plan*. Technical Report v2.0. European Commission, Directorate-General for Research & Innovation. https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf.
- [12] Sebastian S. Feger, Paweł W. Wozniak, Lars Lischke, and Albrecht Schmidt. 2020. ‘Yes, i comply!’: Motivations and practices around research data management and reuse across scientific fields. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), Article 141, 26 pages. <https://doi.org/10.1145/3415212>
- [13] T. Grandon Gill and Alan R. Hevner. 2013. A fitness-utility model for design science research. *ACM Transactions on Management Information Systems* 4, 2 (Aug. 2013), Article 5, 24 pages. <https://doi.org/10.1145/2499962.2499963>
- [14] Arofan Gregory. 2013. *Data Management Planning and the Data Documentation Initiative (DDI)*. Technical Report Version 1.0. Data Documentation Initiative. <https://ddialliance.org/publication/data-management-planning-and-the-data-documentation-initiative-ddi>.

- [15] Marjan Grootveld, Ellen Leenarts, Sarah Jones, Emilie Hermans, and Eliane Fankhauser. 2018. OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans. Retrieved November 9, 2021 from <https://zenodo.org/record/1120245#.YYqm52DMini>.
- [16] Kathryn A. Harrison, Daniel G. Wright, and Philip Trembath. 2017. Implementation of a workflow for publishing citeable environmental data: Successes, challenges and opportunities from a data centre perspective. *International Journal on Digital Libraries* 18, 2 (June 2017), 133–143. <https://doi.org/10.1007/s00799-016-0175-5>
- [17] Rob Hooft, Mateusz Kuzak, Marek Suchanek, and Robert Pergl. 2019. “Data stewardship wizard”: A tool bringing together researchers, data stewards, and data experts around data management planning. *Data Science Journal* 18, 1 (2019), 59. <https://doi.org/10.5334/dsj-2019-059>
- [18] Sarah Jones. 2011. *How to Develop a Data Management and Sharing Plan*. DCC How-to Guides. Edinburgh, Scotland: Digital Curation Center. <http://www.dcc.ac.uk/resources/how-guides>.
- [19] Sarah Jones, Robert Pergl, Rob Hooft, Tomasz Miksa, Robert Samors, Judit Ungvari, Rowena I. Davis, and Tina Lee. 2020. Data management planning: How requirements and solutions are beginning to converge. *Data Intelligence* 2, 1-2 (2020), 208–219. https://doi.org/10.1162/dint_a_00043
- [20] Hamad Rauf Khan, Hsia-Ching Chang, and Jeonghyun Kim. 2018. Unfolding research data services: An information architecture perspective. In *Proceedings of the 18th ACM/IEEE Joint Conference on Digital Libraries (JCDL’18)*. ACM, New York, NY, 353–354. <https://doi.org/10.1145/3197026.3203887>
- [21] Yen-Hsien Lee, Paul Jen-Hwa Hu, and Ching-Yi Tu. 2015. Ontology-based mapping for automated document management: A concept-based technique for word mismatch and ambiguity problems in document clustering. *ACM Transactions on Management Information Systems* 6, 1 (March 2015), Article 4, 22 pages. <https://doi.org/10.1145/2688488>
- [22] Vincenzo Maltese and Fausto Giunchiglia. 2016. Search and analytics challenges in digital libraries and archives. *Journal of Data and Information Quality* 7, 3 (Aug. 2016), Article 10, 3 pages. <https://doi.org/10.1145/2939377>
- [23] William K. Michener. 2015. Ten simple rules for creating a good data management plan. *PLOS Computational Biology* 11, 10 (Oct. 2015), e1004525. <https://doi.org/10.1371/journal.pcbi.1004525>
- [24] Tomasz Miksa and Kevin Ashley. 2017. Workshop report—Research Data Lifecycle and Machine-Actionable Data Management Plans. Retrieved November 9, 2021 from <https://zenodo.org/record/1067753#.YYqoRmDMini>.
- [25] Tomasz Miksa, João Cardoso, and José Luis Borbinha. 2018. Framing the scope of the common data model for machine-actionable data management plans. In *Proceedings of the IEEE International Conference on Big Data (Big Data’18)*. 2733–2742. <https://doi.org/10.1109/BigData.2018.8622618>
- [26] Tomasz Miksa, Peter Neish, Paul Walk, and Andreas Rauber. 2018. Defining requirements for machine-actionable data management plans. In *Proceedings of the 15th International Conference on Digital Preservation (iPRES’18)*. <https://hdl.handle.net/11353/10.923628>
- [27] Tomasz Miksa, Andreas Rauber, Raman Ganguly, and Paolo Budroni. 2017. Information integration for machine actionable data management plans. *International Journal of Digital Curation* 12, 1 (Sept. 2017), 22–35. <https://doi.org/10.2218/ijdc.v12i1.529>
- [28] Tomasz Miksa, Stephanie Simms, Daniel Mietchen, and Sarah Jones. 2019. Ten principles for machine-actionable data management plans. *PLOS Computational Biology* 15, 3 (March 2019), e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>
- [29] Tomasz Miksa, Paul Walk, and Peter Neish. 2020. RDA DMP Common Standard for Machine-actionable Data Management Plans. Retrieved November 9, 2021 from <https://doi.org/10.15497/rda00039>
- [30] Dimitar Nikolov and Esen Tuna. 2019. A lightweight framework for research data management. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (PEARC’19)*. ACM, New York, NY, Article 90, 4 pages. <https://doi.org/10.1145/3332186.3333157>
- [31] NSF Directorate for Engineering. 2018. Data Management Plans Guidance for Principal Investigators. Accessed March 14, 2020 from https://nsf.gov/eng/general/ENG_DMP_Policy.pdf.
- [32] Jay F. Nunamaker Jr. and Robert O. Briggs. 2012. Toward a broader vision for information systems. *ACM Transactions on Management Information Systems* 2, 4 (Jan. 2012), Article 20, 12 pages. <https://doi.org/10.1145/2070710.2070711>
- [33] Simon Oblasser. 2019. Machine-Actionable DMP Application (DMap). Retrieved November 9, 2021 from <https://doi.org/10.5281/zenodo.3522247>
- [34] Simon Oblasser. 2020. Designing an Architecture for Machine-Actionable Research Data Management Planning in an Institutional Context. Retrieved November 9, 2021 from <http://hdl.handle.net/20.500.12708/1103>
- [35] Simon Oblasser. 2020. Oblassers/dmap-mockups: DMap Mockups. Version 1.2. Retrieved November 9, 2021 from <https://doi.org/10.5281/zenodo.3630375>
- [36] Simon Oblasser and Tomasz Miksa. 2019. BPMN Processes for Machine-Actionable DMPs. Retrieved November 9, 2021 from <https://doi.org/10.5281/zenodo.2607556>
- [37] Simon Oblasser, Tomasz Miksa, and Asanobu Kitamoto. 2020. Finding a repository with the help of machine-actionable DMPs: Opportunities and challenges. *International Journal of Digital Curation* 15, 1 (2020), 1–11. <https://doi.org/10.2218/ijdc.v15i1.704>

- [38] John O'Donoghue and John Herbert. 2012. Data management within mHealth environments: Patient sensors, mobile devices, and databases. *Journal of Data and Information Quality* 4, 1 (Oct. 2012), Article 5, 20 pages. <https://doi.org/10.1145/2378016.2378021>
- [39] RDA. 2017. *WG DMP Common Standards Case Statement*. Technical Report. Research Data Alliance. <https://www.rd-alliance.org/groups/dmp-common-standards-wg>.
- [40] Science Europe. 2021. *Practical Guide to the International Alignment of Research Data Management (Extended Edition)*. Technical Report. Science Europe. https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf.
- [41] Stephanie Simms and Sarah Jones. 2017. Next-generation data management plans: Global, machine-actionable, FAIR. *International Journal of Digital Curation* 12, 1 (Sept. 2017), 36–45. <https://doi.org/10.2218/ijdc.v12i1.513>
- [42] Stephanie Simms, Sarah Jones, Daniel Mietchen, and Tomasz Miksa. 2017. Machine-actionable data management plans (maDMPs). *Research Ideas and Outcomes* 3 (April 2017), e13086. <https://doi.org/10.3897/rio.3.e13086>
- [43] Stephanie Simms, Sarah Jones, Tomasz Miksa, Daniel Mietchen, Natasha Simons, and Kathryn Unsworth. 2018. A landscape survey of ActiveDMPs. *International Journal of Digital Curation* 13, 1 (Dec. 2018), 204–214. <https://doi.org/10.2218/ijdc.v13i1.629>
- [44] Stephanie Simms, Marisa Strong, Sarah Jones, and Marta Ribeiro. 2016. The future of data management planning: Tools, policies, and players. *International Journal of Digital Curation* 11, 1 (Oct. 2016), 208–217. <https://doi.org/10.2218/ijdc.v11i1.413>
- [45] Ed Simons, Mijke Jetten, Maaïke Messelink, Marnix van Berchum, Hans Schoonbrood, and Marion Wittenberg. 2017. The important role of CRIS's for registering and archiving research data. The RDS-project at Radboud University (The Netherlands) in cooperation with data-archive DANS. *Procedia Computer Science* 106, C (April 2017), 321–328. <https://doi.org/10.1016/j.procs.2017.03.031>
- [46] Nicholas Smale, Kathryn Unsworth, Gareth Denyer, and Daniel Barr. 2020. The history, advocacy and efficacy of data management plans. *International Journal of Digital Curation* 15 (2020), 1–30. <https://doi.org/10.2218/ijdc.v15i1.525>
- [47] Alexander Tuzhilin. 2011. Knowledge management revisited: Old dogs, new tricks. *ACM Transactions on Management Information Systems* 2, 3 (Oct. 2011), Article 13, 11 pages. <https://doi.org/10.1145/2019618.2019619>
- [48] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3 (March 2016), 160018. <https://doi.org/10.1038/sdata.2016.18>

Received February 2021; revised July 2021; accepted October 2021