

# Active Data Management Planning: chances and challenges

Felix Engel  
FernUniversität in Hagen,  
Germany  
Felix.Engel@FernUni-  
Hagen.de

Heike Görzig  
FernUniversität in Hagen,  
Germany  
Heike.Goerzig@FernUni-  
Hagen.de

Rob Baxter  
EPCC, University of  
Edinburgh  
r.baxter@epcc.ed.ac.uk

Helen Graves  
British Geological Survey,  
United Kingdom  
hmg@bgs.ac.uk

Simon Waddington  
King's College London  
United Kingdom  
simon.waddington@kcl.a  
c.uk

Matthias Hemmje  
FernUniversität in Hagen,  
Germany  
Matthias.Hemmje@fern  
Uni-Hagen.de

## ABSTRACT

There is an increasingly urgent need to ensure the fullest possible preservation of research findings, both for proper validation of the research in question and to enable its comprehensive reuse. Furthermore, research undertakings are expensive and the return on investment needs to be secured by research funding agencies and public funding bodies through proper management of the knowledge that is required for the effective long-term reuse of results. Awareness of these facts leads to increasingly stringent regulations from funding agencies, which seek to enforce compliance of research data with specific policies. Hence, funding agencies are beginning to make *Data Management Plans* (DMP) increasingly mandatory before they will fund a research undertaking. In general, a DMP is a full text document that elaborates how the research data is handled, both during and after the project lifetime. In fact, a DMP includes policies for data management on different levels, as e.g. required by a formal framework the research has to comply with, as well as managerial parameters or policies that directly address the data management system level. Nevertheless, besides the pure establishment of policies, funders and researchers have further requirements concerning active aspects of data management, as e.g.: the continuous adoption of DMPs and its execution throughout the research lifecycle or to preserve the knowledge created during research that is needed for comprehensive later research validation. Because of the complexity of these and further requirements, management support is under discussion in various research communities. Within the international Research Data Alliance, these aspects are discussed within the so called Active DMP Interest Group (ADMP IG, cf. [1]).

This workshop will consider the outcomes of the next RDA ADMP IG workshop (cf. [2]) to discuss additional ADMP related topics and will address further open research questions around ADMPs, with a special focus on continuous adoption of DMPs and automation support. Hence, the aim of this workshop is to identify on base of the submitted contributions and the conclusion of the discussion during the workshop the recent obstacles that

prevent the realization of ADMPs and how those could be addressed. The outcome of this workshop will be the preparation of a roadmap towards a reference framework of ADMP management and automation.

## KEYWORDS

Active Data Management Planning, Policies.

## 1. SUBJECT

*“Innovation has been placed at the heart of the EU's strategy to create growth and jobs”*. As a consequence the European Commission (EC) has encouraged EU countries to invest 3% of their GDP in R&D (cf. [3]). Nevertheless, numerous recent investigations point out that proper validation of research is not a matter of course. The benefits of expensive research are uncertain, because proper reuse is not free of doubt.

To secure the return on investment research funding agencies and public funding bodies are beginning to make proper data management planning of funded research a mandatory part within a project proposal. In general, a DMP is a full text document that describes how the research data is handled both during and after the project. The creation and management of these documents is technically supported by means of services like *DMPOnline* that is offered through the Digital Curation Centre (cf. [4]) or the *DMPTool* that is offered through the California Digital Library (cf. [5]). These services provide funder-specific DMP templates, guiding a user through all the required DMP declarations. But besides the specification of a DMP as a document, further requirements exist from funder and data manager perspective to consider dynamic aspects of a DMP, as a “living document” that is updated continuously and supports automating the policy enforcement while the project progresses. In fact, these living documents would be greatly supported through the existence of tools that would help the researchers in providing the additional information required throughout the lifecycle of the data. This need arises especially, because at the proposal stage some elements of the metadata and other information relating to the datasets are

limited, as e.g. data formats or likely data volumes. Additionally, important it is to enable as well the comprehensive validation of DMP compliance, the monitoring of managed research data or to capture further knowledge that comes into existence along the research undertaking, required for comprehensive later research reuse. Several of mentioned aspects are already under discussion within the RDA, as so called *Active DMPs* (ADMP, cf. [1]). But, while service offers exists to support the creation and management of DMPs, further research is necessary to support as well these active aspects of data management and execution. An initial investigation towards the automation of ADMPs has been investigated in the RDA Practical Policy WG (cf. [6]). This WG discusses so called *actionable rules* that are derived from existing DMPs, formalized and enforced through the application of the *integrated Rule Oriented Data System* (iRODS, cf. [7]) rule language.

## 2. SCOPE

Various actors participate in the process of DMP specification and execution, influencing research data management throughout the whole data lifecycle in various dimensions (cf. [8]). The **Formal Dimension** of DMPs is covered by the funding agencies' Grant Agreements (GA), corresponding legal requirements. The GAs usually provide the contractual framework for DMP; it specifies what the DMP has to accomplish and to comply with. Corresponding laws and regulations therefore provide the legal, regulatory and consequent policy building framework. To comply with the requirements and challenges created by the analysis of this *formal layer*, an RDM work plan is developed in the **Managerial Dimension** of DMP. The RDM work plan describes the RDM scenario that has to be created in order to comply with the DMP requirements and challenges and their corresponding representation schema set up by the analysis of the formal dimension. This RDM work plan includes strategic and organizational aspects, concrete activities, and deliverables. In the RDM work plan sequences of activities and their dependencies are formulated; thus the implementation of the DMP is based on this RDM work plan. The data producers who are, e.g., software developers and researchers on the project, form the **Operative Dimension** of the DMP. Tasks and activities listed in the work plan are executed by them, thereby producing and using the data to be archived and preserved for effective later reuse. Effectively, these three-dimension details the various stakeholders and requirement that are part of the DMP creation and are involved in its development and execution through its lifecycle. But, while the initial creation of DMPs is already supported by means of existing software applications, to date, the consideration and realization of dynamic aspects, spanning all involved actors in data management planning and its execution are rarely addressed within recent discussion and research. Hence, to get a better understanding about the incidents that actually prevent the consideration of the dynamics nature of DMPs, this workshop aims to discuss on the one hand those aspects that actually prevent dynamics in the management and realization of DMPs and on the other hand how these aspects could be addressed through the provision of

software applications that enable automation in data management planning and DMP realization process.

## 3. PROGRAMME STRAND AND CONTENT

The overall scope and goal of the workshop is to bring together academic and industrial researchers and practitioners to discuss a common roadmap to support the acceleration of ADMP-related research activities and to achieve a common understanding of the overall requirements. This roadmap can be used to inform, influence and disseminate ideas to funders, the wider research community, and the general public. Thus, this workshop will maximize the benefit of DMPs for a range of stakeholders. To address this objective, the workshop will bring up and discuss open research questions around active data management planning and automated execution, with a special focus on supporting the treatment of DMPs as living documents, updated and automated where possible throughout the single life cycle phases.

The workshop will start with the presentation of accepted contributions and will be followed by a discussion about identified open issues and preparation of a roadmap that will address the realization of a reference framework towards ADMP management and automation.

## 4. ORGANISATIONAL

**Workshop chair:** Matthias Hemmje

**Co-chairs:** Felix Engel, Heike Görzig, Simon Waddington, Helen Graves.

## 5. REFERENCES

- [1] RDA, ADMP IG. URL: <https://rd-alliance.org/group/active-data-management-plans/case-statement/active-data-management-plans-ig.html> [Accessed 27.042016]
- [2] RDA Workshop on ADMP. URL: <https://indico.cern.ch/event/520120/> [Accessed 27.042016]
- [3] EUROPA - Topics of the European Union - Research & innovation. [Online]. Available: <http://europa.eu/pol/rd/>. [Accessed: 17-Mar-2016].
- [4] DMPTool. URL: <https://dmptool.org/> [Accessed: 22-Mar-2016]
- [5] DMPOnline. URL: <https://dmponline.dcc.ac.uk> [Accessed: 22-Mar-2016]
- [6] <https://rd-alliance.org/groups/practical-policy-wg.html> [Accessed: 22-Mar-2016]
- [7] Chen, S. Y., Conway, M., Crabtree, J., Lee, C., Misra, S., Moore, R. W., . . . Xu, H. (2015). Policy Workbook -iRODS 4.0. Chapel Hill, North Carolina, USA: iRODS Consortium.
- [8] Görzig, H., Engel, F., Brocks, H., Vogel, T., & Hemmje, H. (2015). Towards Data Management Planning Support for Research Data. iPRES2015. Chapel Hill, USA.