

Data Stewardship Profile- Results from a survey of 6 Austrian research-performing institutions

Author: Stefan Reichmann, ORCID: <https://orcid.org/0000-0003-1544-5064>

Institute of Interactive Systems and Data Science
Handlungsfeld Forschung – Programm Digitale TU Graz
TU Graz

DOI: <https://doi.org/10.25365/phaidra.242>



Dieses Werk und dessen Inhalt steht - mit Ausnahme der Logos - unter der Creative Commons Namensnennung 4.0 International Lizenz. Um eine Kopie dieser Lizenz zu sehen, besuchen Sie bitte [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/)

Table of contents

Introduction.....	3
Approach and Sample Description.....	3
Research Questions and Analysis Procedure	4
Results	5
Institutional trajectories: Centralized and decentralized approaches to Data Stewardship and Research Data Management (RDM).....	5
Survey results: Semi-structured questions on data stewardship.....	6
Profile of a Data Steward: What Austrian Universities Look For	6
Tasks of a Data Steward: What are the challenges?	6
Commonalities and Differences	7
Future Implementation of Data Stewardship	8
Summary and Outlook.....	8

Abstract

Universities are beginning to recognize the value inherent in the vast amounts of research data they hold. Yet, universities have different philosophies to tackle the challenges inherent in making the most of these data, depending on their institutional trajectory, aims, and organizational structure. On the road towards curating and stewarding those data, universities encounter a variety of problems. A much-discussed solution involves professionalizing data stewardship at various levels of university organisation. Here, we present the results of an Austrian case study based on a small survey of 6 Austrian research-performing institutions regarding their expectations from data stewards. The survey was developed as part of the FAIR Data Austria project (2020-2022) with the aim of better understanding the desired profile of data stewards in terms of backgrounds and skills in relation to expected task areas. The survey was sent to project partners in preparation for a workshop (held in October 2021) to collaboratively develop the profile of a data steward and find out about project partners current status, expectations, and needs. The survey was targeted at those currently responsible for data management and stewardship (not necessarily under that title) and found that participant institutions find themselves in different places with respect to data stewardship implementation. There are considerable differences in expectations, both in terms of desired backgrounds as well as envisioned tasks of data stewards. This suggests that the profile of a data steward will be somewhat flexible, depending on the kind of institution. In addition, universities find themselves in different places along the way of implementing adequate stewardship mechanisms. The survey found centralized approaches to data stewardship where data agendas lie with central organizational units such as the library, IT services, or (in rare cases) dedicated data working groups as well as the beginnings of decentralized approaches with more specialization.

Introduction

This report summarizes the main findings of a small survey of partners and associated partners in the BMBWF-funded project FAIR Data Austria (forschungsdaten.at). In preparation of a workshop featuring representatives of all partner institutions, TU Graz, as workshop host, conducted a small survey of all partners to gather input on their current trajectories, expectations and plans regarding data stewardship as a possible route towards FAIR data management. The survey yielded six semi-structured questionnaires which were sent to the participant institutions via the project web platform used for collaboration (coLAB, hosted by TU Wien). All six questionnaires contained detailed information on expected competencies and backgrounds of data stewards, their prospective tasks as well as the universities' current strategies. All materials were coded for further analysis. Universities in Austria come from different backgrounds with respect to data management which is reflected in their expectations and approaches. This briefing document summarizes the main findings.

Approach and Sample Description

The aim of this survey was to find out what Austrian research-performing institutions currently associate with the role of data stewards, both in terms of future expectations (what data stewards ought to do) as well as the present situation (where these tasks are currently located and who fulfils them). For this preliminary survey, 6 Austrian universities (all part of the FAIR Data Austria project consortium) have been asked to fill in a short, semi-structured questionnaire about what they consider

current and future tasks and roles of data stewards. The aim was to find commonalities and differences at the institutional level (where possible, taking into account varying institutional trajectories and developments) to come up with a consolidated (vocational) profile for data stewards. The guiding assumption of the survey was that with the data deluge¹, the functions that potential future data stewards are expected to fulfil are already part of someone's job description. The intention therefore was twofold:

- (1) Find out what universities expect from data stewards (Data Steward Models)
- (2) Find out who currently fulfils these functions (Trajectories)

The survey was developed with this idea of a tension between expectations (what is) and realities (what ought). The survey questions were developed accordingly:

- Data Stewardship Profile
 - Tasks of Data Stewards
 - Expectations Data Stewards should fulfil
 - (Soft)Skills
 - Non-Tasks
- Current State: Which tasks of (future) data stewards are currently being managed?
 - Who fulfils these tasks?
 - Where are these people located?
 - What education do these people have?
- Needs
 - What are the most pressing needs in terms of data stewards?
 - What are the advantages and disadvantages of the current model?

It should be noted, therefore, that the primary aim of the survey was to flesh out a preliminary profile of a data steward and what the partaking institutions think should be part of a data steward's job description. The length of the survey did not permit to inquire into the reasons for the prevalence of certain models/preferences. Such a study would have to consider e.g., institutional and historical trajectories of infrastructures, people, and practices (to say the least).

Research Questions and Analysis Procedure

The survey yielded a small sample of 6 semi-structured questionnaires consisting of open questions that were filled in by 6 research-performing institutions across Austria with varying aims (i.e., Full University, Technical University, Medical University and University of Music and Performing Arts). The answers given vary in length and the amount of detail given. In total, the survey yielded about 10 type-written pages of text that was analysed using qualitative data analysis following the approach championed by Timmermans/Tavory² who proposed that qualitative analysis should enable researchers to make discoveries within qualitative material (thus offering an alternative approach both

¹ Borgman, Christine L. (2012). The Conundrum of Sharing Research Data. In: *Journal of the American Society for Information Science and Technology* 63(6), p. 1059.

² Timmermans, Frans; Iddo Tavory (2014). *Abductive Analysis. Theorizing Qualitative Research*. Chicago, London: The University of Chicago Press.

to the more inductive style of grounded theory reasoning) as well as deductive research styles. For our present purposes, suffice it to note that there are already many data stewardship models around that the present analysis can build on that summarize practical experiences to date. In that sense, the approach taken here can be regarded as “abductive” in the sense introduced above, as the questionnaire reflected experiences with and prior research on the topic, with the analysis focused on issues absent from the published literature. In terms of analysing the material, heavy use was made of in-vivo coding (i.e., developing concepts to describe the material, where possible taken directly from the material) and theoretical memoing. These codes were then condensed into categories based on the initial survey questions for easy visualization. In category construction the objective was to learn about commonalities and differences between the 6 institutions surveyed as well as the gap - from the respondents’ perspective - between the current state and their expectations. The construction of the categories allows for relatively simple visualization of these commonalities and differences between expectations and current state. Further, an analysis based on in-vivo-codes allows dimensions of variation to emerge from the empirical material. I.e., based on the analysis we can describe those dimensions of data stewardship models worth looking into. Visualising the data yields a map of commonalities and differences between the institutions on various aspects of the data stewardship profile.

The procedure yielded 135 codes in total, i.e., 135 more or less concrete descriptions of aspects of data stewardship that respondents find relevant. The number of codings associated with each code varies between 1 (mentioned once by one institution) to 6. However, given the open-ended approach to the questions this does not mean that the respective question was answered by all institutions.

Results

Institutional trajectories: Centralized and decentralized approaches to Data Stewardship and Research Data Management (RDM)

An important aspect of the data collected here pertains to where a given institution falls on the spectrum of (possible) approaches to Research Data Management (RDM). In very generic terms, Data Stewardship can be decentralized in the sense that a given institution does not host or employ a central organizational unit to bundle RDM agendas. Practically, this approach plays out in several ways, depending on where RDM tasks are then predominantly located. In some cases, these tasks are situated within institutes (e.g., in the form of an IT specialist who works on RDM agendas some of the time) or in groups (e.g., in the form of - usually senior - scientists who are informally tasked with overseeing RDM). Currently, the six institutions surveyed are spread across the spectrum. To make a cautious attempt at generalization, based on this data alone, RDM agendas are frequently located at the institute/group level, but there are definite attempts to offer centralized support. Where these centralized support structures are located depends on institutional trajectories and resources, among other things. The degree to which RDM is centralized/decentralized also depends on the history of the issue at the respective institution, including past efforts made to establish the topic as well as respective infrastructures. Whether RDM responsibilities are located at the group level depends, inter alia, on the discipline in question (types, amounts, complexity of data, disciplinary culture, etc.).

Education and backgrounds form a second important aspect of this dimension. Depending on where an institution is positioned on this centralized-decentralized spectrum, data stewards' tasks, and consequently backgrounds and desired qualifications, will be more broadly or narrowly defined, with clear implications for the profile of data stewards (e.g. how much IT knowledge is required/desired) as well as the perception of data stewards (e.g., whether their job is regarded as a research service - as part of research services and support - or as part of research - data stewards based within research groups/institutes). In terms of gaining additional information on the approaches taken by the various institutions, an analysis of RDM policies (where they exist) might provide additional insight. In terms of the dimension discussed here, the existence of a centralized RDM structure points to an institutional interest in RDM (in most cases) - as opposed to decentralized, circumscribed approaches that require a lot more initiative on the part of researchers.

Survey results: Semi-structured questions on data stewardship

Profile of a Data Steward: What Austrian Universities Look For

In terms of the profile of prospective data stewards, informants distinguish between knowledge and background of potential candidates as opposed to the tasks that data stewards are expected to fulfil (the distinction is not always clear-cut, however). The skills universities look for in data stewards are diverse and differ considerably across institutions, which suggests - as will be discussed in more detail below - that universities subscribe to vastly different visions when it comes to data stewardship. This is particularly relevant when it comes to certain soft skills such as communication and presentation skills which are clearly more relevant when data stewards' tasks include networking within and outside their institutional context. The issue of philosophies also pertains to skills such as data analysis - to what extent do research institutions demand people with data-analytic skills or data science backgrounds? At present, this seems to be a salient yet unsolved problem, not least because research institutions differ with respect to the extent that data stewards would be involved in operational data management. Interestingly, there is more agreement on the didactic aspects of data stewardship, and consequently on the associated pedagogical skills which suggests that development and delivery of training constitutes a fundamental aspect of a data steward's job profile. At present, there seems to be a tension between domain-specific and generic agendas which is reflected in the desired backgrounds.

Tasks of a Data Steward: What are the challenges?

The institutions we surveyed express unanimity with respect to a large number of data steward tasks. To gauge variance between institutions, we included questions on non-tasks of data stewards as well and found divergent opinions with respect to certain aspects of data stewardship. In some instances, these institutional discrepancies can be explained with respect to institutional structures or their historical trajectories and how these impinge on the way respondents tend to associate certain tasks/roles with certain organisational units. Respondents concur that consulting is a fundamental aspect of a data steward's job profile, even though depending on the institutional set-up the objects of consultation will differ considerably. One such point of disagreement concerns consulting data stewards in matters of proposal writing. In most instances, researchers can solicit advice in writing and

submitting proposals at a dedicated organisational unit which bundles legal expertise with expertise in research funding to offer consultation and training for researchers. As certain aspects of data management are relevant at the proposal stage (writing a DMP, for example), participants expect there to be considerable overlap in competencies, at least at the outset. This holds for centralized data stewards, as embedded data stewards are expected to be much more specialized and hence working operationally within research teams/faculty on day-to-day data management consulting. Repository management is another such area of dissent and is probably even more closely linked to the 3 data stewardship models discussed. In at least one of the institutions under study, the work that is now underway regarding data stewards developed out of earlier efforts to establish an institutional repository. There, data stewardship agendas are handled by a centralized organisational unit with close ties to the repository. However, this may not be a model for other universities where repositories will most likely be hosted by the IT department. The respective areas of divergence are discussed in more detail in the section on embedded versus generic data stewards below.

Commonalities and Differences

The institutions surveyed are found at very different stages in the process of institutionalizing data stewards. The (admittedly very general) dimensions centralized versus decentralized can be found in all institutions, depending (among other factors) on institution size, (historical) conditions, priorities, and expectations. All 3 broad current approaches have distinct advantages. Flexibility was associated with a project-based approach to data stewardship. Centralized data stewardship allows for an overview and, at times, bundling of RDM agendas in one place. Information exchange and mutual learning was described as a major advantage of small, agile approaches to data stewardship (with the caveat that this was only stated by institutions without full-time data stewards, i.e., where data stewardship tasks are part of a broader job description). Implicit knowledge has both advantages and disadvantages. On the one hand, informal contacts with researchers through current, grown infrastructures are regarded as advantageous as they provide centralized data managers with insight otherwise difficult to obtain. On the other hand, informal structures and implicit RDM knowledge/procedures in a variety of places (institutes, research groups) means a lot of diversity and the need to assess said knowledge. Where there is a possibility to access implicit knowledge/procedures, respondents felt that this would lead to researcher-led development of RDM tools and services (and thus to higher acceptance). On the other hand, this means that a centralized RDM structure also tends to lack depth. Trust was mentioned by one respondent as a tremendous advantage of the historical centralized data stewardship model, as well as high levels of commitment of the key actors (data managers).

A major disadvantage of the centralized approach seems to be a lack of structure due to the diversity of RDM practices and a lack of staff. In a centralized model, resources tend to be scarce and don't allow for the development of specialized, discipline-specific knowledge/resources with implications for service quality or overlapping knowledge management structures. On the part of researchers, this means a lack of clear contact points for specific issues, unavailability of discipline-specific consultation and a lack of transparency of service provision. There is a subtle difference within the centralized model of project-based and global funding for data stewardship. Respondents with project-based DS models lament lack of future calculability and unclear sustainability of service provision, as well as a lack of decentralized support within faculties. On the side of education, respondents lament a lack of clear career paths or training for data stewards. Smaller institutions lament that RDM resources are usually split across diverse agendas.

Future Implementation of Data Stewardship

One of the more striking findings concerns the different philosophies with respect to Data Stewardship that can be found across institutions. While there do seem to already be many commonalities between institutions (to be further discussed) depending on their overall structure, aims, and trajectories, there are some remaining issues with respect to data stewardship profiles that concern future tasks and non-tasks. Here, we observed a striking incongruence of expectations across institutions, which entails that there is still some discussion as concerns the profile of data stewards. This is remarkable in the sense that the level of tasks seems to be fairly straightforward at first glance. However, research institutions seem to fall into several groups with respect to what are considered tasks of data stewards (or: based on the materials, we distinguish 3 types of data stewards³). The tasks on which respondents are split are consulting on funding/legal compliance, public relations, project management, consulting in proposal writing, and repository management. Inclusion/exclusion of some of these tasks might change the profile (and required skill set) of data stewards substantially.

Summary and Outlook

The survey was conceived as a preliminary gauge of the current status of data stewardship at Austrian research-performing institutions to preface the Data Stewardship workshop organized by the FAIR Data Austria project and held in Graz on 21 October 2020. The results show research cultures only beginning to attune to the needs of data stewardship that, as a consequence, exhibits a multitude of approaches, models, and expectations regarding data stewards. While there is a lot of overlap in terms of expected backgrounds and tasks of data stewards, there are also areas of considerable dispute, the most salient of which concern the questions of outreach, consulting (where this involves specialized knowledge of a certain kind, such as legal expertise, many aspects of consulting are already taken up by dedicated organizational units such as research services), and repository management. The last is an example of the institutional affordances of data stewards and clearly shows that their implementation will deeply depend on an institution's previous trajectory with respect to data management.

³ S. Scholtens, M. Jetten, J. Böhmer, C. Staiger, I. Slouwerhof, M. van der Geest, C.W.G van Gelder (2019, October 5). Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise. Zenodo. <http://doi.org/10.5281/zenodo.3471708>