LIBER QUARTERLY
The Journal of the Association of European Research Libraries
Vol. 29 (2019) xx–xx | e-ISSN: 2213-056X

Research Data Management and the Evolutions of Scholarship: Policy, Infrastructure and Data Literacy at KU Leuven

Tom Willaert

Vrije Universiteit Brussel, and KU Leuven, Belgium
tom@ai.vub.ac.be, orcid.org/0000-0002-3879-6767

Jacob Cottyn

KU Leuven, Belgium
jacob.cottyn@kuleuven.be

Ulrike Kenens

KU Leuven
ulrike.kenens@kuleuven.be

Thomas Vandendriessche

KU Leuven, Belgium
thomas.vandendriessche@kuleuven.be, orcid.org/0000-0001-9980-6941

Demmy Verbeke

KU Leuven, Belgium
demmy.verbeke@kuleuven.be, orcid.org/0000-0002-1020-3659

Roxanne Wyns

KU Leuven, Belgium
roxanne.wyns@kuleuven.be

Uopen Journals | http://liberquarterly.eu/ | DOI: 10.18352/lq.20272

Abstract

This case study critically examines ongoing developments in contemporary scholarship through the lens of research data management support at KU Leuven, and KU Leuven Libraries in particular. By means of case-based examples, current initiatives for fostering sound scientific work and scholarship are considered in three associated domains: support for policy-making, the development of research infrastructures, and digital literacy training for students, scientists and scholars. It is outlined how KU Leuven Libraries collaborates with partner services in order to contribute to KU Leuven's research data management support network. Particular attention is devoted to the innovations that facilitate such collaborations. These accounts of initial experiences form the basis for a reflection on best practices and pitfalls, and foreground a number of pertinent challenges facing the domain of research data management, including matters of scalability, technology acceptance and adoption, and methods for effectively gauging and communicating the manifold transformations of science and scholarship.

Keywords: research data management; digital scholarship; science policy; infrastructure; data literacy

1. Research Data and Resource Practices

As a much debated topic in recent literature, definitions of what constitute research data abound. Arguably, the most effective of these do not frame research data as a fixed ontological category, but rather relate the concept of data explicitly to the manifold and ever-changing strategies and practices through which researchers interact with resources. Lisa Gitelman and Virginia Jackson, for instance, characterize data as products of interpretation and various other operations of knowledge production, rather than as self-evident, *a priori* information:

> At first glance, data are apparently before the fact: they are the starting point for what we know, who we are, and how we communicate. This shared sense of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself. If we're not careful, in other words, our zeal for more and more data can become a faith in their neutrality and autonomy,

> their objectivity. Think about the ways people talk and write about data. Data are familiarly 'collected,' 'entered,' 'compiled,' 'stored,' 'processed,' 'mined,' and 'interpreted.' Less obvious are the ways in which the final term in this sequence – interpretation – haunts its predecessors. At a certain level the collection and management of data may be said to presuppose interpretation. 'Data [do] not just exist,' Lev Manovich explains, they have to be 'generated.' Data need to be imagined *as* data to exist and function as such, and imagination of data entails an interpretive base. (Gitelman & Jackson, 2013, p. 2–3).

Intrinsically tied to both research practices as well as the notion of resources, the concept of research data foregrounds two interconnected evolutions in contemporary scholarship: a series of methodological shifts emphasizing data-driven and potentially 'open' scholarship, as well as a rethinking of the services required to manage the (digital and physical) resources underpinning these evolutions. As such, recent years have seen the continuation and transformation of scholarly paradigms through notions of among others digital humanities, eScience, artificial intelligence, big data, machine learning and advanced analytics, and, coinciding with these developments, the demarcation of the field of research data management.

Not unlike the concept of 'research data' itself, the term 'research data management' (RDM) is marked by a range of interpretations and definitions. According to the Digital Curation Center (DCC), research data management is mainly paramount to 'digital curation,' involving 'maintaining, preserving and adding value to digital data throughout its lifecycle' (DCC, 2019). Other institutions embrace a more holistic definition that also encompasses the curation of physical data (including for instance tissue samples, archaeological artefacts, rare manuscripts, or geological samples). In their different guises and interpretations, infrastructures and services associated with proper research data management address the entire or specific parts of the data life cycle, that is, the process through which a researcher progresses from developing a management strategy for the research data towards an implemented, sustainable business plan for working with active data, as well as storing, archiving and cataloguing data upon completion of a research project.

The envisaged scientific, economic and ethical gains of proper research data management have been articulated by scientists and scholars across the

sciences and humanities, university administrators and research funders alike (see Krishnamurthy & Awazu, 2016). From the perspective of research integrity, a sound RDM strategy opens up data to scrutiny and criticism, facilitating the growth of knowledge and making researchers more accountable for their findings. Similarly, proper research data management and traceable data can lead to more targeted research efforts, for instance by preventing the collection or creation of duplicate data. Overall, these benefits have been clustered around the prominent acronym of FAIR data: data that are findable, accessible, interoperable and reusable (Mons, 2018; Wilkinson et al., 2016). These principles open up a window onto a scholarly domain in which human researchers work alongside machines (algorithms, information systems) and in which data are both human interpretable and machine-actionable.

The emergence of theories and practices for research data management underbuilds a contemporary knowledge culture marked by debates concerning the openness and transparency of knowledge and research practices. Many European universities have for instance declared their support for open scholarship and science and have thus supported initiatives in the field of open data and more open, inclusive modes of scholarly communication (primarily open access). Importantly, the emergence of open paradigms has been framed as a cultural transition as much as a technical one. As outlined in the recent LERU roadmap *Open Science and its role in universities*, open scholarship

> […] requires a culture change in the way stakeholders in the research, education and knowledge exchange communities create, store and deliver the outputs of their activity (LERU, 2018, p. 3).

Similarly, the eight pillars of Open Science identified by the European Commission (notably the future of scholarly publishing, FAIR data, the European Open Science Cloud, education and skills, reward and incentives, next-generation metrics, research integrity and citizen science) promote a multifaceted, participatory culture surrounding research data that transcends the boundaries of academia (European Commission, 2018; LERU, 2018).

The current discourses, practices and services surrounding research data and RDM thus not only frame data as a scientifically and economically coveted

resource ('the new gold,' 'the new oil,' '(renewable) energy'), but also, as it is considered in the present case study, as a *shared* resource. That is, a resource co-created, transformed and managed by numerous stakeholders. It is in this regard that research libraries play a key collaborative role in this domain. Research data management, as a series of practices for dealing with resources, indeed foregrounds some of the expertise and functions which libraries have held since antiquity, and which they have perfected since they, as Jeffrey Schnapp and Matthew Battles have it, moved 'beyond the book' to embrace (digital) technologies (Schnapp & Battles, 2014). Already at the library of Alexandria, for instance, the challenges of creating, transferring and preserving knowledge were faced by theoreticians and practitioners alike, as ideas conceptualized by philosophers and scholars were inscribed or copied onto scrolls and classified by means of a shelving system, with the aim of opening up knowledge to others. Under one roof, gaps were thus bridged between knowledge management strategies such as the creation of ontologies and taxonomies, and knowledge management practices including scholarly primitives such as the physical acts of writing and copying. As a result of the digitalization and library automation efforts of recent decades, research libraries can operationalize the extended guises of these early resource practices, such as ontology engineering, metadata assignment, data structuring, information retrieval, and an intensified involvement in the research life cycle through efforts in scholarly communication.

This is not to say, however, that library participation in research data management comes without its proper challenges. Support for RDM initiatives on the levels of policy-making, infrastructure development and digital literacy training requires continuous innovations of existing services and infrastructures. The present case study explores these dynamics based on the process through which KU Leuven Libraries equipped itself for durable participation in the changing scholarly domain. To this end, the case study offers a series of comparative analyses that relate KU Leuven's RDM initiatives to theoretical reflections from the relevant literature, as well as to ongoing practical developments at other (European) research institutions. Notably, Section 2 of the case delves into concrete examples of how RDM services were developed in the context of the 2018 call for proposals of the Research Foundation Flanders (FWO). Section 3 then builds on these experiences to identify and elaborate on pressing challenges that require further work in this continuously evolving space.

2. Research Data Management Support at KU Leuven Libraries

KU Leuven Libraries is an organization that consists of 24 libraries and learning centres across Flanders. Through a series of talent development projects for staff in 2014–2015, KU Leuven Libraries, together with the Research Coordination Office, has been involved in the early conceptualization of services and infrastructures to face the challenges of research data management. To consolidate this involvement, support for research data management has since been included as a concrete point of action in the libraries' strategic plans, first for the period 2015–2017, and then – even more explicitly – in the strategic plan for 2018–2021.

A lot of practical information and inspiration was drawn from mapping best practices from other universities, such as TUDelft and UGent, and consequently fine-tuned to meet the needs at KU Leuven. Since 2016, RDM support at KU Leuven is organized around a central RDM support desk, located in KU Leuven Libraries, and currently staffed by two employees. In order to provide sufficient first-line support for RDM, this central RDM support desk is complemented by an interdisciplinary network of subject-specific data experts and data stewards. The data experts consist of a centralized group of approximately five library staff members that have backgrounds in research and research support, and who have the responsibility of staying informed about (international) developments in the field of RDM, participating in the university-wide working groups concerning RDM, and acting as intermediaries between the central support desk and the data stewards. These data stewards are a group of approximately twenty-five information specialists and reference librarians. Embedded within various faculties and libraries across KU Leuven, the data stewards are positioned closely to researchers and are thus best placed to act as first-line support. A key point in this regard was that, until now, no new staff were hired, but that investments were made to retrain staff already on board. The network of data experts and stewards collaborates with the university's IT services, the Legal Department, KU Leuven Research and Development, the Research Coordination Office, and the library systems developer LIBIS.

Researchers can contact the central support desk through a central e-mail address and (since 2019) corresponding ticketing system. Inquiries concerning data types, meta data and documentation, repositories and frequently

asked questions are handled by the support desk in collaboration with the KU Leuven Libraries' network of data experts and data stewards. Apart from questions that are centred around core library services (data types, metadata, repositories), topics of a more specialized nature, such as those concerning data storage and ethical or legal aspects of research, are delegated to the responsible units at KU Leuven. Depending on the topic, the other units either take complete control of the issue raised, or feed back to the support desk, which then communicates this input to the researchers.

In 2017–2018, an internal assessment of the state of digital scholarship services was made at KU Leuven using the RISE framework (Rans & White, 2017). This tool guides research institutions through the process of assessing their research infrastructures by grading them on a scale of one to three, each number indicating increasing degrees of maturity of the infrastructure. The topics covered by the framework are as follows: RDM Policy & strategy, Business plans and sustainability, Advisory Services, Training, Data management planning, Active data management, Appraisal and risk assessment, Preservation, and Access, publishing and discovery. Some of these topics are further divided into a number of (more concrete) sub-topics. E.g. the topic of business plans and sustainability covers the subtopics Staff investment, Technology investment and Cost modelling. One of the major gains of using this framework is not only that it offers a state of the art, but also helps to map the many partners and aspects of research data management support. The final report, prepared by KU Leuven Libraries, IT services and the Research Coordination Office, was presented to the vice rector for research policy in March 2018. It led to the establishment in the summer of 2018 of a steering committee on RDM (with underlying working groups on policy, infrastructure, and advice and training). KU Leuven Libraries, naturally, is one of the units which play an active role in both the steering group and the working groups.

In order to identify some of the challenges and required innovations associated with RDM, and to illustrate the working of the expert network, this section provides case examples from the RDM support that was provided at KU Leuven in the context of the 2018 call for proposals of the Research Foundation Flanders (FWO). Though some services had already been provided earlier in the context of a limited number of Horizon2020 applications, this was the first occasion during which a major funder of research in Flanders required applicants to include an outline for a data management plan, which

made the need for RDM services all the more urgent. The slimmed-down plan required by the FWO consisted of five questions (FWO, n.d.):

1. Describe the datatypes (surveys, sequences, manuscripts, objects, …) the research will collect and/or generate and/or (re)use. (max. 700 characters)
2. Are the following provisions in place in order to preserve the data during and at least 5 years after the end of the research? Please motivate your answer. (max. 700 characters)
    1. Designation of responsible person (if already designated, please fill in his/her name)
    2. Storage capacity/repository
        1. During the research
        2. After the research
3. Is there a reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)
4. If issues concerning research data are indicated in the ethics questionnaire of this application form, will those data require specific security measures; if yes, can they be put in place? (max. 700 characters)
5. Are there other issues related to the data management you think relevant to mention? (max. 700 characters)

Researchers who were awarded the grant were required to submit a fully-fledged data management plan (DMP) within six months after the start of their project. In this plan, various aspects of data management had to be addressed in more concrete detail. The initial five questions however served as a first encounter with research data management. The questions might appear simple at first sight, but they cover and in some cases reframe or defamiliarize fundamental aspects of researchers' interactions with their resources. In this regard, answering these questions required that researchers were introduced to prevailing data policies, appropriate infrastructures and the necessary digital literacy skills. Correspondingly, it was required that the university (and with it the aforementioned expert network) provided support in each of these domains.

2.1. Policies

For researchers, answering the five questions listed above was potentially a first encounter with the policies and regulations surrounding research data. If anything, the prominence of the questions on the grant proposal form was a signal that research data management was high on the agenda of the funder and policy maker they were applying with (the Flemish government). However, the five questions also inevitably bring into scope other top-down regulations that are simultaneously at play on various levels. A researcher's data might for instance be directly or indirectly subject to European data privacy laws (GDPR), data policies at the university, and binding agreements with third parties or publishers. Complex interactions between these different regulations and policies can emerge, for instance when regulations supersede each other in a cascading fashion. Not to be underestimated in this regard are also the many unwritten rules, best practices and conventions for dealing with research data and information that govern disciplines or even individual departments.

At KU Leuven, the current research data policy was first implemented in December 2014. This policy stipulates that in the course of a project research data should at all times be made available upon request to KU Leuven and that the data are to be retained for a period of at least five years after publication or after the end date of the research project grant agreement.[1] However, third-party agreements overrule the university guidelines. Through its policies, KU Leuven explicitly recognizes the fundamental role of research data in high-quality scholarship and scientific integrity. This is exemplified by the fact that all researchers who obtain substantial internal funding (over €600.000) are required to fill out a full data management plan within six months after the start of their project. This way, the university's policy ties in seamlessly with the data management requirements as set out by Europe's Horizon2020 programme and the aforementioned FWO. Anticipating that the requirement of filling in a DMP will extend to other application procedures, not only at a postdoctoral but also at pre-doctoral level, data management plans are already a required milestone in the doctoral training of certain Faculties or Departments at the university, such as the KU Leuven Faculty of Arts.

In order to assist KU Leuven researchers in complying with the various data management strategies, the KU Leuven Libraries RDM expert network

is closely collaborating with various services within KU Leuven. As such, the network fulfils an interfacing role between researchers and specialists in various aspects of research data regulations and policies. Firstly, the Research Coordination Office ensures funder compliance and liaises with the various Ethical Commitees. Secondly, the Legal Department offers tailored advice on the processing of personal data under GDPR and answers questions surrounding ownership. Similarly, the IT services provide technological support for various data types. Third, the Technology Transfer Office (KU Leuven Research & Development) manages agreements with third parties. The expert network's intermediating role between researchers and specialists not only benefits the researchers themselves, but also allows policy specialists to keep track of developments within the research community. The network has for instance helped to procure researcher's feedback on a ledger for personal data as being developed and implemented by the Legal Department.

From a research and policy perspective, research data management support thus provides KU Leuven Libraries with an opportunity to continue developing user-oriented services. For example, when assisting in the writing of data management plans (consisting of five questions or in more elaborate form), the data stewards become actively involved in planning and shaping a research project. In addition, the data stewards gain a valuable insight into the research projects that are being set up at the departments they service, for instance by answering researchers' specific questions or by conducting local surveys on data management practices. Similarly, these interactions allow the network to provide valuable feedback on the policies that are developed on a central KU Leuven level. Finally, members of the expertise network were also acknowledged as partners in research, which for instance follows from the fact that some members of the network have already been cited as co-authors of scientific publications.

2.2. Infrastructures

Contemporary research data management relies heavily on technology in the form of hardware and software that allows researchers to perform a series of operations on their (meta)data, including storing it, sharing it and preserving it for future use or reference. Developing such infrastructures is nontrivial, as it requires experts in the domains of information management and IT to reconcile the research life cycles of different disciplines with a range

of potentially suitable technologies. This translation exercise can be difficult, as it relies on both the researchers' understanding of their (often invisible) technological environments, as well as the IT-experts' awareness of various data life cycles.

A tension at play in this process is that between technologically accommodating existing practices and data management practices, or offering alternative solutions that might methodologically boost ongoing and future research (Gradmann et al., 2016). In both cases, however, knowledge needs to be obtained about researchers' current data management practices and the life cycle of their research data. Data management plans can offer some systematic insight into this matter, but to achieve a wider coverage, a survey for researchers is being developed by the KU Leuven RDM infrastructure working group. This survey gauges among others the types of data that are being gathered, storage practices and requirements, and needs for data sharing.

Through involvement in the KU Leuven working group for infrastructures, KU Leuven Libraries thus also contributes to an environment in which researchers are offered opportunities to be involved in the development of research data infrastructures.

2.3. Digital Literacy and Education

As a series of emerging, technologically-oriented practices for dealing with resources, research data management requires intensive knowledge transfer and training for data stewards and researchers alike. In the start-up phase, the expert network therefore invested in building interpersonal connections and developing training materials that would allow for a smooth exchange of knowledge and best practices among researchers and staff. From the onset, the didactic philosophy adopted by the expert network was one of 'train the trainer.' A heterogeneous group of information specialists and reference librarians was trained so that they could provide first-line support for researchers and staff. In the context of the 2018 FWO call for project proposals, the data stewards became the face of RDM.

Data stewards were familiarized with RDM through information sessions catered to data stewards from the humanities on the one hand, and

biomedical sciences and science and technology on the other. The data stewards were given a brief introduction to the present day scholarly landscape, funding requirements and benefits of research data management. Next, they were given specific information on how the five questions of the FWO could be answered. Data stewards were not provided with ready-made answers or templates, but rather given a basis for further 'on the job'-learning. Questions submitted through the support desk where sent to a subject-specific data steward, each time with a data expert in CC, so that the answers could be fine-tuned and data experts and data stewards could learn from each other. This type of training proved to be an effective way of training a heterogeneous group of data stewards that came from different backgrounds, already took up a number of other tasks and responsibilities, and had varying degrees of familiarity with research methods and practices.

The data experts not only provided internal training within KU Leuven Libraries, but also supported other training initiatives at KU Leuven, with a specific focus on data literacy. This included an optional workshop for PhD students in which the basic principles of RDM are explained and illustrated by the library's data experts, an optional workshop on peer reviewing DMPs that is taught by data experts and members of the research support office, an optional session on publishing data and research funding provided by the research coordination office and a mandatory session on research integrity that also features basic guidelines on research data management.

As was mentioned before, research data management training has also been integrated as part of the Doctoral Schools of the Biomedical Sciences and Science and Technology groups, and the doctoral training programme at the Faculty of Arts. Data experts from KU Leuven Libraries play a central role in this subject-specific RDM training, thus making sure that researchers in the early stages of their career are not only trained in RDM, but also know who they can turn to when they encounter questions or problems. What is more, research data management, and the model of the data life cycle in particular have proven to be a solid base to build new curricula, especially in emerging interdisciplinary fields such as digital humanities. At KU Leuven's Faculty of Arts, for instance, the data life cycle was used as an instrument to structure doctoral training in digital humanities as well as a more fundamental undergraduate course on data and information literacy.

Finally, the staff behind the RDM support desk and other members of the RDM expert network are frequently asked to contribute to external (training) activities, such as RDM workshops organised by the professional organisation for Flemish libraries and archives (VVBAD) or the annual conference *Open Belgium.*

3. Discussion: Practices, Pitfalls and Challenges

The start-up phase of RDM support activities outlined above puts certain theoretical aspects of research data management in a new, practical perspective. Notable practical considerations for implementing RDM support that warrant further discussion are the scalability of these services, fostering technology acceptance and keeping track of the manifold transformations of science and scholarship (both within the organization and beyond).

3.1. Scalability

A key consideration in the implementation of research data management support is its scope and scalability. As a form of research rationalization, research data management is associated with methodological boosting of research domains. In this regard, it can be argued that no one solution or service fits all. Research life cycles can vary greatly from one discipline to another, and developments such as the adoption of digital technologies is moving at different speeds. Similarly, it should be taken into account that there is no such thing as 'raw data.' As outlined earlier, the notion of 'data' itself is intrinsically tied to the researcher's objectives, certain biases that might be at stake, etc. Corresponding notions such as reproducibility, replicability, and 'openness' might be interpreted differently or subject to diverging requirements or restrictions for different disciplines or even individual departments.

The process of developing research data management support thus brings into scope some fundamental questions concerning the nature and practices of science and scholarship. Indeed, a necessary condition for any service to be effective is that it is clearly understood what exactly should be supported. Implementing proper RDM infrastructures and services thus requires an in-depth study and understanding of both the commonalities

and particularities of different modes of conducting science and scholarship. Mapping and describing research life cycles and scholarly domains is a process that requires a balance between on the one hand top-down, theoretically-inspired models and bottom-up, practical input from 'the field.' Scholarly domain models already available in the literature can be of service, although these still need to be reconciled with the actual situation of the organizations, departments or research groups where RDM is to be implemented.

At an organization with a high level of administrative complexity and a wide interdisciplinary scope such as KU Leuven, one challenge for RDM support is to strike a balance between developing common, generic services that can help large groups of researchers, and allowing for the necessary flexibility for tailoring those services to more local needs. Based on their past experiences in the fields that they service (either as researchers or information professionals), data stewards and members of the RDM expert network could provide input for mapping the requirements of various fields. To obtain a broader input at KU Leuven, all relevant units and stakeholders of the university are represented in the aforementioned RDM steering committee and the underlying working groups.

3.2. Technology Acceptance and Adoption

The prominence of research data management marks a series of cultural and methodological shifts that are largely centred around technological innovations. Among these can be counted the digital nature of many data and resources (both born-digital and as the result of digitization), as well as the emerging systems and infrastructures for electronically storing, archiving, sharing and analysing data and information. The possibilities and limitations of these innovations have been well-established and continue to be explored intensively (Borgman, 2010, 2016). Digital data formats and the World Wide Web have for instance opened new possibilities in open access publishing, while at the same time the engineering of paywalls continues to prevent researchers' access to scholarship.

The adoption of research data management technologies and associated practices by researchers is not always self-evident. For some researchers, the investment or learning curve required to optimize their research cycle

might outweigh the perceived benefits. This lack of technology or methodology acceptance can have many causes. On a pragmatic level, research data management requirements can be perceived as additional administrative burdens for researchers and departments that are already subject to budgetary and time constraints. On a more fundamental level, research data management services and infrastructures will encounter resistance when frameworks and methods are not perceived as possible extensions of the research methods and practices in a given field or domain. In turn, there can be many causes for this perceived mismatch. Especially during first encounters with RDM, the question of semantics cannot be underestimated. One of the key impediments that still seem to hinder many implementations of RDM, in particular within the humanities, is the perception that 'data' is synonymous with 'digital data' (and thus associated with the hard sciences). This can be at odds with fields that revolve around notions of 'source material' or 'corpora': concepts that reflect different practices for dealing with resources. In order to be effective, the concept of 'research data' should be explicitly related to these interpretations, and definitions should clearly reflect the fact that data are not a given, but rather created and that there is no such thing as 'raw data' (Gitelman, 2013). Correspondingly, an overemphasis on the technological or digital aspects of 'data' might result in researchers opting out of research data management for fear that they have no 'data' that fits this technologically-driven interpretation.

A pitfall that could be encountered in this regard is that the focus on generalizing research data life cycles might reduce science and scholarship to mechanical (predictable) endeavours. To be truly effective, research data management planning and support should acknowledge the exploratory and oftentimes meandering nature of scientific inquiry. Planning tools and infrastructures should offer sufficient flexibility so that researchers are not restricted in their ability to tailor their research to unexpected intermediary results and promising new approaches. In this regard, it should be made clear that striving for machine-actionable data does not necessarily entail that all human creativity is extracted from the research process, nor that scholarship can or should be left to algorithms. Specifically, the discipline-specific origins of the FAIR principles should be approached critically, and a lot of work remains to be done on how to translate these principles to domains that work with for instance more unstructured data. This observation is also reflected in the experiences at Leiden University:

> The experiences at Leiden University suggest that the success and the impact of data management policies depend, to a large extent, on the availability of ancillary activities that can promote the acceptance of such data management guidelines (Verhaar, Schoots, Sesink, & Frederiks, 2017).

3.3. Gauging and Communicating the Manifold Transformations of Science and Scholarship

Research data management is to an extent an exercise in ontology modelling, that is, in developing a conceptual framework and shared vocabulary that can make research processes intelligible and interoperable. One of the implicit goals of RDM, especially in an open science context, is thus to foster communication of data and information, not only among researchers, but also between academia and the general public. This exchange of data and information can for instance be achieved through standardized data models, metadata vocabularies, and guidelines for writing sound documentation.

From a research data management support perspective, it is beneficial that those who develop and foster this communication obtain sufficient expertise to gauge the needs of their patrons. That is, it is required that data experts and data stewards learn to speak the language of the researchers as well as that of other support partners involved so that the process of service development can happen on shared terms. Similarly, it should be ensured that the types of support provided for different researchers are attuned to each other, as this can increase the interoperability of data and research processes. It can therefore be helpful if those responsible for developing RDM infrastructure have a background in research or are actively involved in certain research practices. However, the extent to which in-depth expertise can or should be operationalized depends on the scope of the services that need to be provided.

A helpful concept to describe the type of expertise that might be required to attune research data management to each other and to the needs of researchers, is Harry Collins's concept of 'interactional expertise.' Interactional expertise is 'the ability to converse expertly about a practical skill or expertise, but without being able to practice it, learned through linguistic socialization among the practitioners' (Collins, 2004, p. 125). As such, interactional expertise is also a medium of communication in scholarship and science:

> The use of interactional expertise go well beyond the sociology of scientific knowledge. For example, all kinds of participant observers, including ethnographers and social anthropologists, at least those who are trying to understand the world of their subjects rather than merely to observe it, must try to acquire it. […] Turning to science and technology, interactional expertise is often the medium of specialist peer review in funding agencies and in journal editing where the reviewers are only sometimes contributors to the narrow speciality being evaluated. It is the medium of interchange within large scale science projects, where again not everyone can be a contributor to everyone else's narrow speciality being evaluated. It is, *a fortiori*, the medium of interchange in properly interdisciplinary, as opposed to multidisciplinary, research. Finally, on those occasions when activists or other concerned persons are driven to it, it can be the medium of interchange between scientists and groups of citizens (Collins, Evans, Robeiro, & Hall, 2006, p. 659).

Those developing or providing RDM support take up a position that is very much like that of a 'participant observer' or peer reviewer (the latter sometimes in the most literal sense). Often not fully versed in the particularities of the research projects they support, they have to engage with researchers in order to come to a common understanding and learn a language that allows them to assess the needs of scholars from different fields. Acquiring this interactional expertise is highly dependent on interactions and engagement with the research community. However, gaining and maintaining sufficient exposure to ongoing developments in scholarship in and of itself can be challenging, as research supporters not only need to take stock of the state of the art at their own institution, but also need to include international dynamics in their scope.

4. Conclusion

Debates surrounding research data management are intrinsically tied to broader dynamics in contemporary science and scholarship. This includes embracing digital technologies, as well as the overall societal reflex to engage more critically with data and information. As illustrated in the present case study, the challenges faced by practitioners open up perspectives to some general challenges facing those fostering research data management, including questions of scalability, technology acceptance, and communication

and expertise. These issues are not new, but the ongoing rationalization and digitization brings them to the forefront. In the face of these ongoing evolutions, it is safe to say that the definitive guide on how to implement RDM services has yet to be written. One could even debate whether or not such a guide can ever be written, as history has seen its share of shifting standards for what constitutes proper science and its associated preferred methods. For the moment, however, an incremental approach in which different domains are methodologically boosted at different speeds seems like a valuable future avenue to explore. Building on shared infrastructures, policies and training modules, RDM service providers need to develop the interactional expertise and shared language required to meet the needs of specific researchers or research groups. As it is impossible to be prepared for every eventuality, knowledge and experience gained 'in the field' will prove of immense value over the coming years.

References

Borgman, C. (2010). *Scholarship in the digital age. Information, infrastructure, and the Internet.* Cambridge, MA: MIT Press.

Borgman, C. (2016). *Big data, little data, no data. Scholarship in the networked world.* Cambridge, MA: MIT Press.

Collins, H. (2004). Interactional expertise as a third kind of knowledge. *Phenomenology and the Cognitive Sciences*, *3*(2), 125–143. https://doi.org/10.1023/B:PHEN.0000040824.89221.1a.

Collins, H., Evans, R., Robeiro, R., & Hall, M. (2006). Experiments with interactional expertise. *Studies in History and Philosophy of Science A*, *37*(4), 656–674. https://doi.org/10.1016/j.shpsa.2006.09.005.

DCC (2019). *What is digital curation?* Retrieved February 4, 2019, from http://www.dcc.ac.uk/digital-curation/what-digital-curation.

European Commission (2018). *The European cloud initiative*. Retrieved February 4, 2019, from https://ec.europa.eu/digital-single-market/en/%20european-cloud-initiative.

FWO (n.d.). *Data Management Plan (DMP*). Retrieved February 4, 2019, from https://www.fwo.be/en/the-fwo/organisation/data-management-plan/.

Gitelman, L. (Ed.) (2013). '*Raw data' is an oxymoron.* Cambridge, MA: MIT Press.

Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Ed.), *'Raw data' is an Oxymoron* (pp. 1–14). Cambridge, MA: MIT Press.

Gradmann, S., Hennicke, S., Tschumpel, G., Dill, K., Thoden, K., Pichler, A., …, Stiller, J. (2016). Modelling the scholarly domain beyond Infrastructure. In: E. Burr (Ed.), *Digital Humanities im deutschsprachigen Raum* (book of abstracts) (pp. 143–146). Leipzig: Universität Leipzig.

Krishnamurthy, R., & Awazu, Y. (2016). Liberating data for public value: The case of Data.gov. *International Journal of Information Management*, *36*(4), 668–672. https://doi.org/10.1016/j.ijinfomgt.2016.03.002.

LERU (2018). *Open Science and its role in universities. A roadmap for cultural change*. LERU Advise paper no. 24. Retrieved February 4, 2019, from https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf.

Mons, B. (2018). *Data stewardship for open science. Implementing FAIR principles.* New York: Chapman and Hall/CRC.

Rans, J., & Whyte, A. (2017). *Using RISE, the Research Infrastructure Self-Evaluation Framework*, v.1.1 Edinburgh: Digital Curation Centre. Retrieved February 4, 2019, from www.dcc.ac.uk/resources/how-guides.

Schnapp, J.T., & Battles, M. (2014). *The library beyond the book.* Cambridge: Harvard University Press.

Verhaar, P., Schoots, F., Sesink, L., & Frederiks, F. (2017). Fostering effective data management practices at Leiden University. *LIBER Quarterly*, *27*(1), 1–22. https://doi.org/10.18352/lq.10185.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., …, Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018, n.p. https://doi.org/10.1038/sdata.2016.18.

Note

[1] A revised research data policy is currently being drafted and will probably be implemented in the course of 2019.