Data Sharing in the High-throughput University: Mobility and Precarity

Stefan Reichmann

Abstract

The nascent field of data science and the expansion of the higher education sector share surprising affinities. The emergence of the “entrepreneurial university” has brought increasing differentiation of the work roles of academics in addition to increased mobility and, for some, precarity. At the same time, researchers are dealing with unprecedented amounts of data. The present article describes how policies and infrastructures implemented to support researchers with data curation tasks might be repurposed by research administrators to tackle problems of academic mobility rooted in increasing precarity of non-tenured research staff. Findings suggest that the organizational benefit of research data management (RDM) is not increased efficiency or reusability of research, but rather increased control over data left behind by non-tenured staff. Recent interest in data mobility needs to be understood by reference to increased researcher mobility. While the view of data as context-independent evidence has been challenged by reference to the investments necessary to mobilize data as evidence in the first place, the material presented here suggests that RDM is repurposed by universities as a strategy to manage, not data, but rather increasing rates of staff turnover. The mobility of data producers and the immobility of data are frequently in tension. Handing over data is problematic irrespective of domain, data type, and funding source. The term “high-throughput university” is introduced in opposition to “high-throughput” data production techniques to suggest that findability and reusability of data need to be recontextualized with reference to increased academic mobility.

Introduction: The Mundane Practice of Research Data Management

The convergence of research and digital information technologies has revolutionized the conduct of research (Finholt and Birnholtz, 2006). Like industry, academia is increasingly keen on extracting value from the vast amounts of (digital) research data it holds (Borgman, 2012;

Leonelli, 2016). Research institutions are expected to implement high-level policies governing research data management (RDM) (Higman and Pinfield, 2015). Formal data archiving and curation processes have been central to certain fields – survey-based social science, climate science, biology – for a long time (Mayernik, 2019). Recently, these practices have been more closely associated with increased transparency (Fecher and Friesike, 2014), better reproducibility, accessibility, sharing, and collaboration (Vicente-Saez and Martinez-Fuentes, 2018). The underlying notion of “data-intensive” research is a recent invention that first appeared in policy initiatives from the early 2000s such as eScience, eInfrastructure, or cyberinfrastructure (Borgman, 2015, p. 31). In promoting the accessibility of scientific outputs, data sharing is fundamental to Open Science (Andreoli-Versbach and Mueller-Langer, 2014; Anagnostou *et al.*, 2015). Publicising research data in (digital) repositories is a precondition for making them accessible and reusable (Piwowar and Vision, 2013). Following increasing recognition of its benefits by the scientific community (McKiernan *et al.*, 2016), Open Science is being widely institutionalized to unfold its intended benefits for scientific knowledge production. This includes policies to increase public access to publicly funded research (Suber, 2012) and to encourage data sharing (Wallis, Rolando and Borgman, 2013; Couture *et al.*, 2018). Accordingly, universities and other research-performing institutions are reorganizing their IT and information support services to accommodate for increased data management and curation needs and to support their researchers in managing their data. Data management and curation is essentially infrastructural work (Bowker 2005) in the sense of being ubiquitous and integral to scientific work but often overlooked (Borgman, 2015). Data curation and related difficulties are a constant preoccupation of researchers working in data-intensive disciplines (Mayernik, 2016), but while data curation is fundamental to achieving reproducibility and reusability, researchers tend to regard it as a mundane occupation, the domain of “technical” specialists at best or of research administrators at worst.

The mundanity of data practices, their marginalized position among researchers, and how they relate to academic mobility and accelerating academic knowledge production (Kim, 2017) are the topics of this article. Over the last three decades, entrepreneurialism emerged as the new paradigm in academic management (Slaughter and Rhoades, 2010), leading to increased institutionalized competition both on the macro- and micro levels. Data are frequently functionally immobile or at least require large amounts of work in order to move (Leonelli, Spichtinger and Prainsack, 2015; Leonelli, 2016). This has to do with knowledge frequently being tacit (Collins, 1985), and it has been suggested that data curation means rendering (tacit) aspects of data practices explicit (Leonelli, 2016). The main claim of this paper is that the stakes in RDM need to be understood with reference to increased researcher mobility: Insofar as ensuring the mobility of data involves explicating (some) tacit dimensions of data practices,

these efforts are potentially repurposed by research administrators to control effects of researcher mobility.

The present study builds on the nascent field of studying data journeys (Leonelli, 2020) as well as the concept of entrepreneurial universities to analyse institutional affordances of RDM implementation and its conflicts with researchers' expectations and needs. It attempts to answer the following questions: What do researchers expect from a large-scale implementation of RDM? What sorts of conflicts does this produce? What agendas is RDM made to serve? The reasons for managing research data in the first place as described by participants are revealing, as they suggest that *the need to manage data is rooted in the fixed-term nature of research projects and the consequent high rate of staff turnover*. The interview data on which this paper is based are drawn from a study of a research institution in Austria, one of the first countries to start investing in Open Science in a coordinated manner (Mayer *et al.*, 2020). Taking the concept of the entrepreneurial university as a theoretical starting point, the article explores how RDM is often unwittingly deployed as a solution to problems rooted in the neoliberal re-organization of the university, i.e. high turn-over of research staff and a lack of common standards to hand over data. It is therefore crucial to understand both universities' motivations and agendas in fostering Open Science and data management, and researchers' stakes in sharing their data (competitive advantage versus access to others' data).

The contribution is structured as follows: Section 2 gives an overview of the OS landscape in Austria and beyond as a background to current RDM initiatives. Section 3 briefly reviews the theoretical background, most notably recent work on data journeys as well as the development of the entrepreneurial university. Section 4 describes the methodology and sample. Section 5 discusses perceptions and observations by informants regarding the implementation of RDM. Section 6 interprets these findings in light of the perspectives introduced in section 3 to argue that beyond considerations of practicability and motivation to manage and share data, an emergent latent function of RDM seems to be to accommodate the upheavals caused by liberalizing the academic job market. In this, the paper hopes to contribute to opening the organizational aspects of RDM practices for analysis by showing that informants interpret RDM as a strategic means of detaching competences and capacities from their bearers to retain institutional control. Section 7 spells out these conclusions and attempts to answer the research questions.

Open Science in Austria and beyond

Over the past decade, research funders, publishers, and policy makers have reacted to the possibilities (and constraints) offered by digital data and scientific computation (Mayernik,

2016). Open Science and data sharing promise to generate increased citation rates (Piwowar, Day and Fridsma, 2007), increase transparency (Gilmore *et al.*, 2017) and reproducibility (Toelch and Ostwald, 2018), foster economic growth (Tennant, Jacques and Collister, 2016) and more efficient science (Leonelli, Spichtinger and Prainsack, 2015). What started with Open Access policies (Suber, 2012) has since become a movement to make the entire research process more transparent and participatory. Notably, many countries have started to develop and implement Open Science policies (Olesk, Kaal and Toom, 2019; Lilja, 2020). Since the early 2000s, Austria has seen a number of initiatives (Mayer *et al.*, 2020). The Austrian Science Fund (FWF) has been promoting Open Access since 2004 and has been a member of cOAlition S since 2018. The Open Science (formerly: Open Access) Network Austria was founded in 2012 by the council of the rectors of Austrian universities¹ and the Austrian Science Fund as a think tank to formulate recommendations for the implementation of Open Science. In 2016, the group published its first recommendations for a national Open Science strategy (OANA/UNIKO 2016). The 12 “Vienna Principles” on the future of scientific publishing were published in the same year² and have served as a framework for strategy and policy development, both nationally and internationally. In addition, 2016 marks the foundation of RepManNet, an exchange forum for Austrian repository managers. Since then, many Austrian research institutions have developed Open Science policies, and some have started to implement institutional repositories. As of 2019, Open Science is part of the HRSM (Structural Funds for Higher Education) focus on digitisation, which is currently funding several projects to develop Open Science infrastructures and coordination between Austrian beneficiaries. (Mayer *et al.*, 2020, p. 18 ff.)

Related Work

Research Data Management and Data Mobility

A vast literature studies data management practices (Tenopir *et al.*, 2011, 2015; Unal *et al.*, 2019), including governance (Grant, 2015), and infrastructures (Knight, 2015; Bugaje and Chowdhury, 2018). Studies show little awareness of RDM (Mancilla *et al.*, 2019), a lack of technical skills (Aydinoglu, Dogan and Taskin, 2017) and missing institutional guidelines (Chigwada, Chiparausha and Kasiroori, 2017). There is staggering variability in practices, infrastructures, approaches (Akers and Doty, 2013) and data types (Borgman, 2015; Whitmire, Boock and Sutton, 2015; Borghi and Van Gulick, 2018), both within and between research domains (Mayernik, 2016). Conceptually, recent work on data management and sharing is

¹ <https://uniko.ac.at/>

² <https://viennaprinciples.org>

unified by the assumption that data are more or less context-free representations of underlying phenomena (Leonelli, 2016, p. v) and that data can, at least in principle, be modelled and interpreted using statistical tools. Data are simultaneously conceived as context-free representations of reality and therefore universals (which accounts for their epistemic role as arbiters of replicability and accountability) and tangible outcomes of research (Leonelli, 2020).

An alternative view holds that it should be possible for data to be publicly scrutinized because data provides evidence for scientific claims (Leonelli 2016: 56). In this way, researchers are construed as bound by scholarly altruism to share their data for the benefit of science (Lilja, 2020, p. 13). Indeed, altruism has been found to be a significant factor in data sharing behaviour (Cho, Chen and Chung, 2010). Borgman's (2015) ethnographies spanning multiple research disciplines similarly claim that data are "entities used as evidence of phenomena for the purposes of research or scholarship" (Borgman, 2015, p. 29) and stresses how both the characteristics as well as the meaning of data change with their context: with the specificity of the research purpose, with the scope of data collection, and with the goal of the research (Borgman, 2012, p. 1063). These views are unified by the functionalist assumption that data is an umbrella term for a variety of objects whose common denominator is their expected evidential value (Mayernik, 2019). Consequently, according to these views, data sharing occurs for a variety of reasons (Borgman, 2012, p. 1066): to reproduce/verify published research, to enable others to ask new questions based on extant data, to make publicly funded research results publicly available, and to advance research (Leonelli, Spichtinger and Prainsack, 2015).

Data sharing, the practice of making data available to the public, is deeply entrenched within some fields such as climate science (Edwards, 2010) and biology (Leonelli, 2016). Looking beyond these examples, however, only a small proportion of data are shared (Leonelli, 2016, p. 57). Indeed, the promotion of open data as one key to advancing basic research is not without resistance, as many researchers fear loss of competitive advantage, voice concerns about intellectual property as well as issues of privacy and trade-offs with time spent disseminating data (which not everywhere reflected in promotion and tenure criteria) (Leonelli, 2016, p. 58). Recently, the emerging research programme of studying *data journeys* has pointed towards the resources implicated in making data (re)usable across institutions, research contexts, fields, and national borders (Leonelli 2020). Empirical work on data mobility (see the excellent collection in Leonelli/Tempini 2020) shows that like other scientific outputs, data are the result of manifold processes of construction (Knorr Cetina, 1991, 1999; Cetina, 2007). In fact, a tremendous amount of work and curation is necessary to make data mobile (Mirowski, 2018), e.g. in terms of developing and applying appropriate metadata (a conception that echoes Latour and Woolgar's (Latour and Woolgar, 1986; Latour, 1994) conception of

laboratory products as immutable mobiles and suggests that data are not mobile in and of themselves (Leonelli, 2016, 2020)).

Metadata: Tacit Knowledge and Data Mobility

How data are implicated in scientific practices is relevant to understanding their (im)mobility. Research activities depend on implicit knowledge that is not easily documented and often shared informally between research sites and research communities (Collins, 1985; Knorr Cetina, 1999). Organizational performance requires effective knowledge management (Baruch, 2013; Lefebvre, Bakhtiari and Spruit, 2020). Tacit knowledge resides with people (von Krogh et al. 2012). Therefore, “people and their management are significant for universities” (Baruch, 2013, p. 197). Making data shareable requires making tacit knowledge explicit (Leonelli 2016), because reusability of data produced by others depends on their interpretability (Kurata, Matsubayashi and Mine, 2017, p. 3). Contextual information is codified using metadata, which are thereby integral to scientific knowledge production (Edwards *et al.*, 2011). Typically, metadata contain information about the content as well as the context of a dataset, “information necessary for investigators separated by space, time, institutions, or disciplinary norms to establish common ground” (Edwards *et al.*, 2011, p. 672). They constitute an explicit attempt to articulate (some) tacit (contextual) dimensions of knowledge (Borgman, 2012; Leonelli, 2016, p. 100). Kurata and colleagues (2017) helpfully distinguish “explicit” context (metadata) from “implicit” context (tacit knowledge) to explain how data always “exist in a context” (Borgman, 2015, p. 4). However, what is deemed “context” and what is deemed “data” or “metadata” is itself the outcome of social processes (Knorr Cetina, 1999) that determine what is useful as evidence (Borgman, 2015; Leonelli, 2016). While falling short of improving the prospects for replication in general, the practice of consulting metadata is likely to improve the work of researchers from the same fields and increase the likelihood of replication (Leonelli, 2016, p. 113).

Data sharing depends on digital infrastructures (Bardi, Casarosa and Manghi, 2018) as digital technologies are needed for data storage, analysis, publication and dissemination (Fecher and Friesike, 2014). The introduction of computer systems potentially reifies conflictual relationships within organizations (Bloomfield, 1989). User-friendly programming “masks a subtle deskilling of the user” (Linn 1985: 87, cited after Bloomfield, 1989, p. 420). Digital information management systems serve a dual role in organizations, simultaneously used by workers to accomplish their tasks and by management to monitor workers’ success (Suchman, 2007, p. 203). As the material presented here seems to show, however, the goals of university management and research staff diverge. While there is an increasing expectation for research institutions to implement RDM governance, Open Science practices are not adopted with open

arms (Lilja, 2020). Overarching aims to foster transparency and reproducibility are not commonly shared (Mayernik, 2019). Institutionally mandated data management (Peng *et al.*, 2018) often operates by centralizing tacit aspects of research skills with university administration to solve problems created e.g. by transitions in the academic employment structure such as increasing rates of staff turnover. I therefore turn to academic employment next.

Higher Education Expansion and Academic Employment

The Emergence of Academic Capitalism

Since the 1980s, universities have started to incorporate economic and social development as part of their mission (Etzkowitz, 1998), with research results increasingly being defined as intellectual property (Etzkowitz and Webster, 1994). Since then, entrepreneurial activities at universities such as university start-ups, university patenting, technology transfer, and university-industry co-authorship have increased significantly (Shibayama, 2012). Through changing the requirements for public research support, governments hoped to use universities “as surrogate agents for industrial policy programs that governments were unwilling to undertake more directly” (Etzkowitz and Webster, 1994, p. 497). At the same time, a wide range of industrial sectors relies on knowledge in order to remain competitive (Cetina, 2007; Etzkowitz and Viale, 2010). Universities increasingly act as producers of knowledge and human resources instead of being restricted to teaching and research (Etzkowitz and Leydesdorff, 2000, p. 110). Entrepreneurialism emerged as the new paradigm for development in higher education institutions in the 1990s (Filippakou and Williams, 2014). Accounts of “academic capitalism” describe the complex social processes of marketization, increasing “managerial” governance, and increasing competition within and between universities (Clark, 1998; Slaughter and Rhoades, 2010; Münch, 2014; Rhoades *et al.*, 2019). Clark (1983) proposed that universities exist in a threefold tension between (academic) norms, state regulation, and market forces. Debates about how best to provide public services led to a shift towards the market boundary, although to varying extents across national academic systems. These changes were most visible e.g. through increases in the percentage of income “from sources other than the government” (Filippakou and Williams, 2014, p. 72). The rise of academic capitalism has led to a loss of autonomy and strong uniformity in terms of research trajectories and institutional forms (Münch, 2014). Entrepreneurialism and the associated increase in competition are responsible for many of the current problems in higher education (Schulze-Cleven *et al.*, 2017), such as difficulties retaining control over research outputs. When looked at from the perspective of the needs of university administrations, the material

presented in this article suggests that mandatory RDM is a strategy deployed to control some of the effects of increased staff turnover.

Academic Mobility and the Emergence of a Globalized Academia

Transnational academic mobility has gained in importance in the wake of academic capitalism (Kim, 2017), strategically promoted by policy actors, e.g. as a criterion for career evaluation. Universities preferably employ senior staff who obtained their qualification elsewhere because “studying abroad is believed to broaden networks and prevent parochial cultures” (Bennion and Locke, 2010, p. 11). The expansion of higher education systems has led to growing pressure on institutions and individuals in terms of qualifications, societal impact (Hicks and Holbrook, 2020), and mobility (Bennion and Locke, 2010). Job market conditions for advanced degree holders have become increasingly insecure (Baruch, 2013) as the availability of permanent academic positions declined (Acker and Haque, 2015, 2017). Fixed-term positions are on the rise and cause increasing career planning difficulties (Münch, 2014). Internationalization is acute in science and technology-related fields (Kim, 2017), while a decline of available academic positions is felt predominantly by SSH graduates (Acker and Haque, 2017). Within the EU, these developments were streamlined under the reforms of the so-called Bologna process, launched by the Bologna Declaration in 1999 and renewed in the Budapest-Vienna declaration in 2010 (Kehm, 2010), which led to increasing complexity of university governance across the signatory countries as the agenda gradually shifted from structural changes to curricular reform. In Austria, the process has had far-reaching consequences for what was formerly open access to tertiary education, though pressures for reform were as much domestic as external (Pechar and Pellert, 2004).

Approach and Methods

Case Selection

This article is based on ethnographic material and interview data. The data for this study were collected at an institution with a science and engineering focus. Quantitative results have been described in detail elsewhere (Reichmann *et al.*, 2020). The context of this research is a long-term project with the aim of developing and implementing RDM infrastructures and policies. As of 2019, the university features approx. 1800 researchers in 97 institutes across seven faculties, 5 (predominantly) engineering, one (predominantly) scientific, and one (predominantly) humanities, with an annual research output of approx. 2300 publications (2019). In total, 18 researchers and staff in research-related roles were identified using

snowball sampling. In addition, ethnographic work was carried out in 2019 that consisted of informal interviews with research administrators to introduce the project, solicit interviews and collect (top-down) perspectives on RDM. As part of this process, a working group consisting of faculty representatives collaborated in developing the RDM policy and to document the perspectives of faculty research staff. The ethnographic work was carried out as part of this process. Interviews average 46 minutes in length. Interview partners were first approached via faculty deans who acted as gatekeepers. Owing to the large variety of data practices especially in the sciences and engineering domains, interviewees were sampled with regards to maximum variation across faculties/institutes. Interviewees were predominantly full professors, associate professors, and PIs from all faculties. Some PIs brought PhDs or Postdocs to the interviews unsolicited to answer questions about data management.

For the most part, RDM has been studied either at the level of research institutions, disciplines, or sub-disciplinary units. Disciplines comprise complex social units with fluid boundaries (Stichweh, 1994). As has been suggested elsewhere (Wallis, Rolando and Borgman, 2013), many issues related to RDM are independent of discipline, either in the sense of transcending disciplines or high internal variability. This has resulted in the assumption that solutions should either be discipline-agnostic or adaptable across fields (Mayernik, 2016), a view that has been criticized for conceiving RDM in predominantly technocratic terms (Finholt and Birnholtz, 2006) and without regard for the fundamental nature of the phenomenon (Kurata, Matsubayashi and Mine, 2017). These studies mostly adopt a domain- or technology-centric perspective on data that risks ignoring the essential tension between economic (e.g. competitive advantage) and epistemic considerations (e.g. reproducibility) (Leonelli, 2016). The present study focuses on the two domains (Ribes, 2019, p. 528 f.) of (natural) science and engineering to avoid these problems. The author's role has been to facilitate and manage the development of RDM policies.

Interview Protocol and Data Collection

All interviews were conducted in German and lasted between 45 and 90 minutes. A semi-structured questionnaire was used, including 8 general questions, each with more open-ended sub-questions. We attempted to gather information about four broad topics:

- a) Respondents' current research focus and role(s), research questions and methods³
- b) Types of data respondents (typically) work with and how data figure in their research process

³ The strategy was for respondents to describe the phenomena they were interested in in their research. The approach is inspired by a classic paper by Bogen and Woodward (1988).

- c) Institutional guidelines for data handling, data management and curation (if applicable), and dissemination of research outputs
- d) Needs and potential areas of improvement

Each interviewee was asked the same set of questions. Follow-up questions were geared to the specific situation and problem definitions. In total, the interviews yielded roughly 700 minutes of recorded material, which was coded using the qualitative analysis software package Rqda (Huang, 2016).

Analysis Procedure

The aim of the study was exploratory. Therefore, I started with an open coding approach to all interview transcripts in search of emergent categories and themes. Initially, coding was done without imposing any theoretical constraints on the coder (as suggested by Glaser and Strauss, 2009). Following this initial approach, small meaningful units of text were coded, either single paragraphs, or, where essential for understanding an interviewee's response, paragraphs including interviewer questions/responses. The abductive analysis approach recommended by Timmermans/Tavory (2012) was employed thereafter, aiming for theoretical innovation (i.e. moving beyond inductive analysis) relative to the vast literature on RDM practices. Analysis was performed on the entire material in three rounds, starting with emergent codes and moving to mature code hierarchies and categories through extensive use of theoretical memoing. Presentation of results is restricted to the conflict between institutional and individual stakes in data sharing and curation. It should be noted that the central claim of this article is owed, in part, to a methodical artefact. Doing a study on research data management entailed asking explicit questions about data management practices, in those words. The term "management" elicited resistance on the part of respondents who across the sample interpreted the term as implying additional administrative burden.

Results

Researchers have mixed feelings regarding RDM. The interviews document ambivalence with respect to streamlining RDM and illustrate how researcher mobility is regarded as a problem in need of an (institutionalized) solution. Some interview excerpts suggest that RDM is interpreted as an attempt by university administration to "order from without" (Suchman, 2007) what are ultimately regarded as matters internal to research. Respondents interpret data management not as a solution to problems of data governance but rather of high rates of turnover for non-tenured staff. Since the latter shoulder the bulk of data analysis tasks,

institutionalizing data management threatens to further undermine their position as it is deployed by the institutions that employ them to ensure that the tangible products of their work remain within the control of the university. On the one hand, researchers interpret data management predominantly as an administrative task (i.e. extra box-ticking without any tangible reward) – thereby drawing a strict boundary between administrative work and research – while at the same time hailing data management as a solution to long-standing problems in research administration – how to render data reusable in the face of increased workforce mobility. Findability (as in FAIR) is thus reframed by respondents as an answer to a problem usually not envisaged by open data advocates.

Social Aspects of Findability: Leaving Data Behind

Who shoulders data curation and analysis tasks? All informants concurred that the bulk of data collection and analysis is performed by PhD students and Post-Docs under conditions of fixed-term employment. Generating data is considered a fundamental aspect of education for (junior) researchers, with implications for the prospects of wide-spread data sharing: “Of course, a lot of the data come from studies by PhDs or Post-Docs, work that ends with a PhD or master thesis [...] our record is then, maybe not a database, but scientific output, our publications” (Associate Professor, Mechanical Engineering and Economic Sciences). Transitioning PhDs or Post-Docs create the problem of migrating data for their successors (field notes, meeting with faculty representatives). Standardisation of (meta)data practices is frequently named by respondents as a solution to this problem. As a corollary, respondents locate RDM training at the level of PhD curricula. Increased mobility of data curators provides a possible explanation as to why RDM is often undervalued. Data curation takes time and having to leave the university after a project ends means that the work invested into RDM will not benefit the person who did it (field notes, meeting with representatives of the Faculty for Electrical Engineering). Tenured researchers tend to assume that this is what keeps data producers from seeing the benefits of RDM. In this sense, the utility of RDM is, strictly speaking, a downstream utility. This has nothing specifically to do with educating PhD students to do proper data management. Few institutes have centralized RDM; PhD students are expected to manage their own data without receiving much guidance, which entails that supervisors develop processes for students to manage and hand over data. Respondents in managerial roles (e.g. faculty deans, institute heads) think about RDM issues first in terms of benefits for administration, and only then for researchers. The pressure on researchers to manage data seems particularly high during the final phases of their PhD:

“Well this is the difficult part, because especially during the final phase of the dissertation or master thesis [students] are under a lot of pressure and stress, which tends to be forgotten and then we are

pressured to provide a grade and then people leave. I have thought about the potential to optimize, because we have lots of data and it works, but it could be better.” (Senior scientist, Technical Chemistry and Process Engineering)

The final phase of a research project constitutes a crucial moment in RDM, as this is usually a phase where data producers already cope with large workloads. These tend to overpower data curation tasks, not least because documentation is rarely considered part of research. Several factors contribute to data management being conceived as an administrative task that predominantly benefits, not the data producers, but their respective successors or those on unlimited contracts:

„This is the first point, long-term archiving and of course, that these things are interoperable, that would be the second point. How do we go about managing all this data centrally so that every project researcher can simply add their data for those who are here long-term and do not leave after [completing] their thesis?” (Full Professor, Technical Chemistry and Process Engineering)

It is evident from the above that data management is conceived as predominantly benefitting the institute, and not necessarily the person that has to do it. Governance is needed to “separate [RDM] from the persons [who perform data curation]” (Senior scientist, Computer Science and Biomedical Engineering). And further:

„[We need] a concerted, focussed effort to build such a database, with corresponding data entry masks, [...] and someone watching over people typing in their things correctly and that they input everything. You put in the work once so that in the long-term [RDM] becomes easier, less time-consuming.” (Full Professor, Technical Chemistry and Process Engineering)

The same is true not just for data but also for code:

“The codes that our colleagues write are collected on some server. There’s enormous potential for improvement in terms of sustainability. [code] is lying around in some PhD student’s file folder, once this person leaves the next one has to start from scratch [...] it’s predominantly PhD candidates that are concerned with code.” (Associate Professor, Mechanical Engineering and Economic Sciences)

This respondent interprets aspects of FAIR data management (interoperability and reproducibility) as attributes of the succession of data producers. A small ad-hoc working group assembled for a presentation of the university’s RDM strategy suggested installing clusters of data management practices instead of developing faculty-specific policies without consulting those affected. The same group suggested that those responsible for populating the repositories should be the first to see the benefits, e.g. increases in reproducibility (fields notes, meeting with representatives of the Faculty for Electrical Engineering). Given the high turnover rates of PhD researchers, though, this entails that institutes need to find ways of dealing with employees leaving behind a trail of data.

“As long as people don’t vanish completely, which has not happened a lot, thank god. You can save the physical data for all eternity, with the guarantee of being preserved forever, but I don’t think highly of that [...] it’s highly inefficient if my young colleagues work on their own and develop [their own ways of managing data]. If I have the chance of bringing them into contact with their predecessor for a half-day or a day, this is one hundred times more effective than doing it alone. (Associate Professor, Mechanical Engineering and Economic Sciences)

Certain aspects of RDM, most notably findability, are frequently interpreted in terms of organizing the succession of data producers. Data management is more readily interpreted in terms of managing staff, not in terms of data curation. Findability is not necessarily, at least in the view of our informants, a result of appropriate data curation, but rather of managing the succession of data producers. This is especially evident when non-digital data are concerned, such as lab notebooks:

“They [lab notebooks] are archived, yes. I don’t have any at the moment, they used to be, we used to have a collection, but they are with the [...] none of my students have finished their theses yet, but this is exactly where they are collected. When the notebooks are nicely done then what has been done can be reproduced. That does work, and like I said, the data are handed over to the supervisor [of the thesis], in the form of USB sticks, say, and the supervisor takes care of them until publications are ready. [...] As concerns the analogue data, the lab journals are handed over, this stays with us as well.” (Senior Scientist, Technical Chemistry and Process Engineering)

(Analogue) Lab notebooks provide an excellent example of what is at stake as there, the administrative aspects of findability are less hidden away by technology. Issues of data management are not constituted by digital technologies and arise irrespective of data type. If the findability of data is an organisational achievement, reproducibility (as an aspect of reusability of results) and replicability are thus outcomes of how well institutes manage to organise the succession of PhD candidates:

„[Reproducibility] is a big problem for us, but this is down to people’s know-how [...] it is not at all easy to pass on the know-how from one PhD candidate to the next, because when one leaves you have a 6-month-break until the next one starts, that is the problem.” (Full Professor, Electrical Engineering)

Discussions of tacit knowledge (e.g. Leonelli 2016; Collins 1985) aside, the problem of reproducing results is – at least in this instance – attributed to the break that ensues between two successive PhD candidates. This has led some informants to trust personal contacts over technical solutions:

“The best data management works via a quick phone call when you say, ‘How did you do that the other day, how was that?’, which tops every documentation” (Associate Professor, Mechanical Engineering and Economic Sciences)

The above excerpt is especially pertinent to the social aspects of findability, and, by extension, interoperability of data (see also Ribes, 2017). Whether “a quick phone call” is appropriate will depend on the data in question, but the excerpt certainly documents a tendency to trust social arrangements more than technical solutions in some instances.

Division of Labour: The Mundanity of Research Data Management

Respondents attribute problems of insufficient data curation (lack of interoperability or findability) to certain aspects of academic division of labour as well as a lack of protocols for managing employee mobility. Respondents therefore interpret (centralized) RDM as a potential solution to these organizational problems. In some instances, collaboration between scientific and technical staff is closer, but this does not preclude scientific staff to shoulder the bulk of data-related tasks:

“Well, doctoral students, each works on their research project [...] industry projects are being supervised by doctoral students as well, but this is not what they do for their PhD research [These projects] have a project lead, a PhD student, who sees to it that things are running smoothly, and there are technicians whose task is to take care of the photos” (Associate Professor, Mechanical Engineering and Economic Sciences)

And further:

“They [PhD students] are employed at the university and the technicians are the interface to building the model. The PhD students do the planning and help the technicians build the model, they have to see to it that it works [...] we have two technicians taking care of the measurements, setting up the equipment, that have to set up the equipment for the [...] measurements, and that usually carry out the measurements, but regrettably they are not the ones doing the data analysis [...] They do the measurements and then hand the data over, which is when the PhD students come back in, as they are the ones who proceed to deliver results based on these data.” (Associate Professor, Mechanical Engineering and Economic Sciences)

Even where technicians are involved in the process of data creation, data analysis and curation ultimately fall to PhD students. In part, this is because creating a unique data set is often considered a fundamental part of receiving a PhD. That non-academic staff is involved in data creation feeds into the view that data management has less to do with good scientific practice than with administrative “box-ticking”:

„Especially as concerns data management, it should be kept simple, because you do see this among non-scientific staff, but also with PhD candidates, they drown in work and resent additional bureaucratic expense” (Associate Professor, Mechanical Engineering and Economic Sciences)

In the above excerpt, data management is squarely associated with bureaucratic expenses (and importantly not with good scientific practice). Outsourcing this work to data stewards is welcomed:

“If that really works with the data steward and if there is a uniform data structure at least for a subsection where you can say, well, now I actually can find my colleague’s PhD candidate’s data easily, I think that would be a good idea [...] most of those that produce data would say that it would be cool if someone took care of [data management]” (Full Professor, Mechanical Engineering and Economic Sciences)

Industry projects often make specific demands on confidentiality and data protection, producing the opposite problem: How to ensure that no-one can find data, even when researchers work with them at least for the duration of the project?

“Well, I think I prefer that when a PhD candidate leaves, they take it with them, back to the company, because even when I make an effort, some PhD candidate will find it years later, ignorant of the original agreement, which is when I have a problem guaranteeing secrecy. I handle the problem like this, everything that is secret stays exclusively with the person working on it, because that way access is restricted automatically. Access can be restricted on the server as well.” (Interview with a Full Professor, Electrical Engineering)

The above excerpt provides a mirror image of the organizational problem of rendering data findable. As the informant explains, the secrecy that is often mandated by third-party research projects can be achieved by restricting data access to the person (or persons) working with data.

Data Sharing and Academic Mobility

Motivations for or against sharing data have been discussed extensively (Campbell *et al.*, 2000; Oushy *et al.*, 2015; Ceci, 2018; Houtkoop *et al.*, 2018). RDM practices and technical possibilities have adjusted where sharing data is regarded as fostering the common interest of a research field:

“This was sort of an early necessity, because only when you had all the data could you, for instance when it comes to identifying stems, well when you have a bacterium, only if you have the database from all of them are you able to compare it and then you can really find out the unknowns, whereas if you only have your own data, the certainty is much, much lower, and that was, I think, that were very smart people and in microbial ecology we had, that was always a community where these things were discussed.” (Full Professor, Technical Chemistry, Chemical and Process Engineering, Biotechnology)

The above excerpt is from an interview with Life Scientists who describe their community as very interactive. The community has recognized the benefits of data sharing for a long time

(Leonelli 2016), which might be an additional reason for the highly developed culture of sharing:

“Very social, very communicative and I think therefore that was a logical development, people saw that it helped the entire community and because there aren’t a whole lot of loners in the communities, there are certain communities where it will certainly take longer and be more difficult, but here people saw the added value and started to work collaboratively” (Full Professor, Technical Chemistry, Chemical and Process Engineering, Biotechnology)

However, not everyone sees added value in making data available. The interviews clearly show that whether a practice is seen as conducive to a common good is not simply a matter of demonstrating its efficacy:

“I mean, in terms of availability, I think, making truly large amounts of data globally available is not, there’s not a lot of demand for me to do that. I think the question is, what is available that will help us or what we have [...] it’s not customary in our community to really make data, raw data available, freely available, at other departments, is very rare” (Associate Professor, Civil Engineering Sciences)

Respondents recognize that mobility of staff may have an impact on the individual willingness to share data and manage them accordingly:

“Okay it’s not mine, it belongs to [the university], and that you accept. But should I really pass that on to my successor? I mean I do it, but don’t like to do it, he should not have it too easy” (Full Professor, Mechanical Engineering and Economic Sciences)

The above excerpt documents ambivalence engrained in data sharing. The tension in the above excerpt is indicative of the conflict of interest between research institutions and researchers implicit in data sharing. In the view of many respondents, data sharing imposes a certain discipline on researchers to develop data management structures that enable proper reuse and data security. As has been argued above, these demands extend to the primary data producers, i.e. PhD students:

“If there is a uniform data structure for one area at least, where you can say, okay, now I can find my colleague’s doctoral candidate’s data easily, that, I think, is a good idea” (Associate Professor, Civil Engineering Sciences)

The required personal discipline associated with data sharing comes with both positive and negative connotations. In the above excerpt, the requirement to make data available is readily interpreted as positive, in the sense of forcing researchers to keep datasets well curated. On the other hand, this discipline is precisely what some perceive as an additional “administrative” burden, which suggests that there is some flexibility in the way respondents interpret research data management.

High-throughput universities: Mobility of researchers, not data

Almost all interviews feature a recurring narrative which holds that increased mobility of the academic workforce due to the expansion of the higher education sector has adverse effects such as knowledge loss (even though the former is desired by university management). While many respondents squarely acknowledge the additional administrative burden associated with RDM, the expectation is that the implied additional workload should be borne by PhD students as well:

“When data amounts increase, when projects develop in such a way that you have to archive more data, handle more data, then answers will develop how to deal with this situation [...] my advice would be to train, not at the level of professors, but at the level of Post-Docs, because they are the ones who handle data. I have nothing to do with that, come to think of it, I might have a look [...] I don't have data, but others do [...] I need to be able to trust that people manage their data well, that once PhD projects are finished, that the data are properly stored and made available. But honestly, everybody knows from experience that the connection from one person to the next never works 100%. Maybe in rare cases, but there will always be a cut when people switch [institutions].” (Full Professor, Mechanical Engineering and Economic Sciences)

While the above excerpt acknowledges that increasing data amounts pose an issue, the last two sentences identify the root problem as handing data over from one person to the next. This situation is pressing for all faculties as everywhere, fixed-term contract researchers shoulder the bulk of data collection. Some respondents in higher administrative roles deem high fluctuation desirable because new staff is believed to bring in new ideas, expressly making a high percentage of (fixed term) PhD positions part of their strategy. The idea of training PhD students through collecting and curating data looms large in many of the interviews:

“In the end, we only have four permanent staff members. [NAME] is the only scientist, in addition to three technical assistants. That's all we have, everybody else is on a fixed-term contract and in training.” (Full Professor, Computer Science and Biomedical Engineering)

Evidently though, employing researchers on fixed-term contracts is associated with problems, the most pressing of which is knowledge loss:

“Our doctoral quota is high [...] two theses per year, which means that I have a lot of fluctuation among my assistants and project assistants, which makes research data management a very pressing problem [...] The biggest problem is that people leave with all their knowledge. We have tried to set up a database and failed most of the time because pdfs are not really searchable [...] data archiving after PhD theses is done unsystematically, which why I hope to get solutions out of this digitalisation of research project” (Full Professor, Mechanical Engineering and Economic Sciences)

High turnover is a social problem recognized by respondents as requiring organizational solutions. An important but often overlooked aspect of RDM is communication between domain experts and between generations of data managers (see e.g. Anderson *et al.*, 2007):

“The successor [of a doctoral student] should start before the first one leaves, and in most cases we manage to do that, to build up the people early enough [...] that the first one can’t say, ‘well there you go, there’s a USB-drive where you’ll find everything.” (Associate Professor, Mechanical Engineering and Economic Sciences)

Switching to a model of employing data specialists e.g. at the faculty to provide support with data management tasks would take this sort of expertise away from researchers. There seem to be different coping mechanisms. One of these is to offer training in data management. In general, the desire for training can, in many cases, be explained by reference to high mobility those members of staff primarily responsible for data management:

“The question is, what are you going to do? You’d have to, right from the start, once a new assistant takes up a position, take a full week to train them on guidelines and data structures” (Associate Professor, Mechanical Engineering and Economic Sciences)

Managing data producers and managing data are thus two sides of the same coin. In those cases where data management is deemed important, it is almost exclusively so for its role in securing data for the institution. In cases where the desired organizational structures are in place, RDM is only of secondary concern. This isn’t always easy, though, because securing data succession through organizational processes requires adequate funds:

“It [fluctuation] is quite annoying, really, and my only way around that is good documentation and, if possible, to have people’s contracts overlap. But this is often impossible due to the financial situation.” (Associate Professor, Mechanical Engineering and Economic Sciences)

The desire for RDM training can be explained by reference to high turnover among those members of staff primarily responsible for data management. The link between archiving data and finding a solution to the problems of employee mobility is particularly salient in the following excerpt:

“This would be the first point, data archiving and of course, that these things are linked. How do you manage to administer these data as centrally as possible and in such a way that every project employee can simply type in their data [...] for example for those people who are here longer and don’t simply vanish after their Master’s thesis.” (Full Professor, Technical Chemistry and Process Engineering)

RDM is sought as a potential remedy for some of the immediate effects of high rates of staff turnover. The problem of making data available for reuse seems to occur predominantly for research that is publicly funded (whether basic or applied; there are exceptions, such as research on issues of security). Recall here that securing the availability of publicly funded

research results constitutes one fundamental rationale for sharing data in the first place (Borgman 2012). The same is not true for data from privately funded research, e.g. from industry projects, because in those instances, funders have an interest to keep data confidential. There is an additional (but overlooked) implication for data management associated with short-term, project-based, industry-funded research. Their short-lived nature entails that the structural problem of handing data over does not seem to be as pressing in those instances:

„Those research topics are partitioned into small project chunks which means that we often succeed in completing a project within the lifespan of one PhD thesis [...] There, the problem of handing over is not as pronounced. (Associate Professor, Mechanical Engineering and Economic Sciences)

This is a solution to the problem of leaving data behind which is not predicated upon technical or social measures, but rather turns on project organisation. Whether handing over data becomes problematic or not therefore depends on external circumstances as well, such as project duration and whether it is possible to answer a research question within the time frame of a research project.

Mobile researchers and immobile data

In this article, I have presented empirical and conceptual work on the affordances of research data management strategies on research organisations and researchers. While it has been found earlier that institutional mandates for research data management will be followed only to the extent that they can accommodate existing norms of practice and institutional constraints (Wilsdon *et al.*, 2015; Curry *et al.*, 2020), the effects of administrative constraints have been studied less frequently. The difficulties with RDM perceived by the researchers in this sample suggest that problems of data reusability and findability are as much social as they are technical. One finding in support of this claim is that the general problem of organising data handovers between data producers and their successors (or their institution) occurs irrespective of domain, funding (public or private), and data type. Respondents report that these issues arise at a crucial moment, not in the data life cycle, but in the life cycle of PhD projects. The life cycle of PhD projects frequently overlaps with that of third party-funded projects, but the two are not always coextensive. This situation typically creates the problem of finding a process of handing data over from the data producers – usually someone working on a fixed-term project, either as a PhD or a Postdoc. Some of the constraints that give rise to the development of data management processes can thereby be traced back to project-based research.

Research is organized in terms of third-party funding across all research domains (critically: Mayernik, 2016). This might be one explanation for the fact that respondents construe some of the (social) aspects of RDM as domain-independent, as does a significant portion of the literature (Mayernik, Batcheller and Borgman, 2011; Wallis, Rolando and Borgman, 2013). On the other hand, intradisciplinary data practices are just as varied as interdisciplinary practices (Borgman, 2012; Wallis, Rolando and Borgman, 2013). As Mauthner and Parry (2013) point out, the way data sharing objectives are formulated in RDM policies tends to obscure these variations in research practices which are context-based, situational, contingent, and pragmatic (Knorr Cetina, 1999). Uneven power relations (Lilja, 2020, p. 10) are an equally important but somewhat less salient issue. Current RDM policies are frequently based on a domain-independent understanding of RDM issues, which is certainly present in the way respondents describe constraints on RDM. While respondents welcome the advantages of more coordinated approaches to RDM (the expectation that dedicated data professionals will free up time for research tasks being a case in point, see Leonelli 2016), they have a clear idea who would benefit and who would shoulder the work. According to the functionalist paradigm of the 1950s, social patterns can be functional at the system level yet dysfunctional at the individual level (Strevens, 2006). Most respondents in this study are not directly involved in data curation and analysis, but squarely agree that as primary data producers, PhDs (and Postdocs) are the ones shouldering the bulk of data curation tasks, even when most of their employment contracts have an expiration date, which creates the need for organized data curation in the first place. The material described in this paper was gathered before any formal RDM policy implementation. The rejection of the aims of formalized RDM that was found is attributable to a predominant interpretation of RDM as an extrinsic administrative agenda, rather than serving the intrinsic improvement of the research process. The amount of curation required to render data shareable was considered by most respondents to be extra work.

As universities are increasingly introducing managerial practices taken over from the corporate world, researchers experience increasing employment insecurity (e.g. Slaughter/Rhoades 2004), but also increasing administrative burden (Bennion and Locke 2010; Münch 2014). I would like to suggest here that efforts to introduce data management should be interpreted against this backdrop of macro-transformations of the university towards a “liberalization” of the academic labour market, with highly mobile data producers and increased rates of staff turnover that necessitate more dedicated data management. If we follow this line of thought, one of the most striking findings concerns the conflictual nature of data practices. Not too long ago, many accounts of data sharing could be criticized for considering it a predominantly technical problem which, given the right infrastructure, was a matter of course (Finholt/Birnholtz 2006). The findings presented here point to more vexed, conflictual motives to oppose research data management. The interviews document ambivalence engrained in

data sharing resp. data retention. Respondents are quite aware that their employment contracts oblige them to hand over their data (the details of this procedure vary) to their employer since per agreement, all data produced by university employees belong to the university. At the same time, respondents have a very clear idea about the potential competitive advantage associated with keeping data secret, at least from one's successor: Where researchers compete on the same turf, there is a clear incentive not to share data. Presumably, this is part of the reason the university sees the need to oblige its employees to do so. The ambivalence with respect to data sharing is indicative of the conflict of interest between research institutions and researchers implicit in data sharing.

Discussion and conclusions: High-throughput institutions and immobile data

Open Science seeks profound cultural change of the scientific system⁴. However, the main arguments to promote Open Science practices may not address the practical barriers (such as existing incentive structures) involved in changing researchers' behaviours (McKiernan *et al.*, 2016). For the Life Sciences, the practice of making data available has been encoded in the FAIR guiding principles (Wilkinson *et al.*, 2016) which mandate that data be Findable, Accessible, Interoperable, and Reusable⁵. Documenting data to meet these standards, the material described here shows, can in many instances not be reduced to making datasets machine-actionable, however. Indeed, research on data practices by Science and Technology Studies (STS) scholars has noted that data sharing comes with steep demands on data curation and management (Leonelli, Spichtinger and Prainsack, 2015; Leonelli, 2016, 2020). Here, I proposed an alternative analysis of data practices in claiming that the transition to Open Science needs to be understood as much by reference to broader social transformations of academia (e.g. with respect to work regimes) as by reference to increasing data amounts.

At the outset of this article, I posed three interrelated questions:

⁴ Though, as Mirowski (2018) has pointed out, it is not always clear what Open Science intends to fix about old science. In Mirowski's view, Open Science is much closer aligned to platform capitalism than to a structural break in scientific practice.

⁵ The 15 principles associated with FAIR data management refer to machine-readable data, as is evidenced by the mandate that data be assigned persistent identifiers, be indexed in a searchable resource, are retrievable by standard protocols, etc. The principles "define characteristics that contemporary data resources, tools, vocabularies and infrastructures should exhibit to assist discovery and reuse by third-parties" (Wilkinson *et al.*, 2016).

- 1) What do researchers in leadership roles versus researchers in non-leadership roles expect from a large-scale implementation of RDM?
- 2) What kinds of conflicts does this produce?
- 3) What agendas is RDM made to serve?

The main conclusion I draw from the material is that data management can be repurposed to cope with the effects of increased staff mobility. The dominant answer to the first question is that all respondents agree that institutional data management structures will serve to solve certain problems inherent in increased staff mobility. Data mobility and researcher mobility are connected insofar as the bulk of research data are produced by temporary staff, which makes it difficult for universities to control what happens to data after an employee's transition to a new employer. Importantly, respondents in leadership roles predominantly had a positive outlook on staff mobility, but not necessarily some of its consequences. This is not to claim that research data management itself has (or could have) adverse effects on data producers; it does mean, however, that some aspects of the Open Science agenda might actually be made to serve goals other than increasing the transparency of the research process. The material presented here suggests, as I have tried to show, that RDM in particular is deployed to govern mobility of researchers, not of data, which suggests that RDM is not just concerned with the consequences of high-throughput data processing techniques (Reuter, Spacek and Snyder, 2015) – the prevalence of which is not restricted to any one research domain – , but about the organisational consequences of high-throughput universities.

It seems that some aspects of the Open Science agenda, most notably Open/FAIR data, are more politicized than they superficially purport to be. Issues surrounding research data management are already highly politicized and therefore cannot be understood solely in terms of technological (or cultural) advancement. Rather, efforts to streamline research data management seem to reify an inherent conflict of interest, even though at the macro level, individual data producers stand to benefit from increased availability of data. The cultural change implied in introducing Open Science needs to run deeper than is often assumed; the factors that (might) inhibit researchers from sharing their data are as much individual career prospects as they are tied to data producers' institutional position. These results are somewhat limited by the fact that no PhD students were asked to share their views on the matter (given that this aim was not part of the original study). However, awareness that RDM has potentially adverse effects on data producers' control over data suggests that there is at least one more barrier to the adoption of Open Science practices besides career prospects and increased effort.

References

- Acker, S. and Haque, E. (2015) 'The struggle to make sense of doctoral study', *Higher Education Research & Development*, 34(2), pp. 229–241. doi: 10.1080/07294360.2014.956699.
- Acker, S. and Haque, E. (2017) 'Left Out in the Academic Field: Doctoral Graduates Deal with a Decade of Disappearing Jobs', *Canadian Journal of Higher Education*, 47(3), pp. 101–119. doi: 10.7202/1043240ar.
- Akers, K. G. and Doty, J. (2013) 'Disciplinary differences in faculty research data management practices and perspectives', *International Journal of Digital Curation*, 8(2), pp. 5–26. doi: 10.2218/ijdc.v8i2.263.
- Anagnostou, P. *et al.* (2015) 'When data sharing gets close to 100%: what human paleogenetics can teach the open science movement.', *PLoS ONE*, 10(3), p. e0121409.
- Anderson, N. R. *et al.* (2007) 'Issues in Biomedical Research Data Management and Analysis: Needs and Barriers', *Journal of the American Medical Informatics Association: JAMIA*, 14(4), pp. 478–488. doi: 10.1197/jamia.M2114.
- Andreoli-Versbach, P. and Mueller-Langer, F. (2014) 'Open access to data: An ideal professed but not practised', *Research Policy*, 43(9), pp. 1621–1633. doi: 10.1016/j.respol.2014.04.008.
- Aydinoglu, A. U., Dogan, G. and Taskin, Z. (2017) 'Research data management in Turkey: perceptions and practices', *Library Hi Tech*, 35(2), pp. 271–289. doi: 10.1108/LHT-11-2016-0134.
- Bardi, A., Casarosa, V. and Manghi, P. (2018) 'The European Project OpenUP: OPENing UP New Methods, Indicators and Tools for Peer Review, Impact Measurement and Dissemination of Research Results', in Serra, G. and Tasso, C. (eds) *Digital Libraries and Multimedia Archives, Ircdl 2018*. Berlin: Springer-Verlag Berlin, pp. 240–248. doi: 10.1007/978-3-319-73165-0_24.
- Baruch, Y. (2013) 'Careers in academe: the academic labour market as an eco-system', *Career Development International*, 18(2), pp. 196–210. doi: 10.1108/CDI-09-2012-0092.
- Bennion, A. and Locke, W. (2010) 'The Early Career Paths and Employment Conditions of the Academic Profession in 17 Countries', *European Review*, 18, pp. 7–33. doi: 10.1017/S1062798709990299.
- Bloomfield, B. P. (1989) 'On Speaking about Computing', *Sociology*, 23(3), pp. 409–426. doi: 10.1177/0038038589023003005.
- Borghi, J. A. and Van Gulick, A. E. (2018) 'Data management and sharing in neuroimaging: Practices and perceptions of MRI researchers', *PloS one*, 13(7), p. e0200562.
- Borgman, C. L. (2012) 'The conundrum of sharing research data', *Journal of the American Society for Information Science and Technology*, 63(6), pp. 1059–1078. doi: 10.1002/asi.22634.
- Borgman, C. L. (2015) *Big data, little data, no data: scholarship in the networked world*. Cambridge, MA, London: The MIT Press.
- Bugaje, M. and Chowdhury, G. (2018) 'Identifying Design Requirements of a User-Centered Research Data Management System', in *Lecture Notes in Computer Science, Information*

Systems and Applications, incl. Internet/Web, and HCI, pp. 335–347. doi: 10.1007/978-3-030-04257-8_35.

Campbell, E. G. *et al.* (2000) 'Data withholding in academic medicine: characteristics of faculty denied access to research results and biomaterials', *Research Policy*, 29(2), pp. 303–312. doi: 10.1016/S0048-7333(99)00068-2.

Ceci, S. J. (2018) 'Scientists' Attitudes toward Data Sharing:', *Science, Technology, & Human Values*. doi: 10.1177/0162243988013001-206.

Cetina, K. K. (2007) 'Culture in global knowledge societies: knowledge cultures and epistemic cultures', *Interdisciplinary Science Reviews*, 32(4), pp. 361–375. doi: 10.1179/030801807X163571.

Chigwada, J., Chiparausha, B. and Kasiroori, J. (2017) 'Research Data Management in Research Institutions in Zimbabwe', *Data Science Journal*, 16. doi: 10.5334/dsj-2017-031.

Cho, H., Chen, M. and Chung, S. (2010) 'Testing an integrative theoretical model of knowledge-sharing behavior in the context of Wikipedia', *Journal of the American Society for Information Science and Technology*, 61(6), pp. 1198–1212. doi: <https://doi.org/10.1002/asi.21316>.

Clark, B. R. (1983) *The higher education system: Academic organization in cross-national perspective*. Berkeley, CA: University of California Press.

Clark, B. R. (1998) 'The entrepreneurial university: Demand and response', *Tertiary Education and Management*, 4(1), pp. 5–16. doi: 10.1007/BF02679392.

Collins, H. M. (1985) *Changing Order. Replication and Induction in Scientific Practice*. 1st edn. London, Beverly Hills, New Delhi: SAGE Publications.

Couture, J. L. *et al.* (2018) 'A funder-imposed data publication requirement seldom inspired data sharing', *PLOS ONE*, 13(7), p. e0199789. doi: 10.1371/journal.pone.0199789.

Curry, S. *et al.* (2020) *The changing role of funders in responsible research assessment: progress, obstacles and the way ahead*. report. Research on Research Institute. doi: 10.6084/m9.figshare.13227914.v1.

Edwards, P. N. (2010) *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*. London, Cambridge, MA: The MIT Press.

Edwards, P. N. *et al.* (2011) 'Science friction: Data, metadata, and collaboration', *Social Studies of Science*, 41(5), pp. 667–690. doi: 10.1177/0306312711413314.

Etzkowitz, H. (1998) 'The norms of entrepreneurial science: cognitive effects of the new university–industry linkages', *Research Policy*, 27(8), pp. 823–833. doi: 10.1016/S0048-7333(98)00093-6.

Etzkowitz, H. and Leydesdorff, L. (2000) 'The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations', *Research Policy*, 29, pp. 109–123.

Etzkowitz, H. and Viale, R. (2010) 'Polyvalent Knowledge and the Entrepreneurial University: A Third Academic Revolution?', *Critical Sociology*, 36(4), pp. 595–609. doi: 10.1177/0896920510365921.

- Etzkowitz, H. and Webster, A. (1994) 'Science as Intellectual Property', *Handbook of Science and Technology Studies*. Thousand Oaks, London, New Delhi: Sage.
- Fecher, B. and Friesike, S. (2014) 'Open Science: One Term, Five Schools of Thought', in *Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Cham, Heidelberg, New York, Dordrecht, London: Springer, pp. 17–48.
- Filippakou, O. and Williams, G. (2014) 'Academic Capitalism and Entrepreneurial Universities as a New Paradigm of "Development"', *Open Review of Educational Research*, 1(1), pp. 70–83. doi: 10.1080/23265507.2014.964645.
- Finholt, T. A. and Birnholtz, J. P. (2006) 'If We Build It, Will They Come? The Cultural Challenges of Cyberinfrastructure Development', in BAINBRIDGE, W. S. and ROCO, M. C. (eds) *Managing nano-bio-info-cogno innovations*. Dordrecht: Springer Netherlands, pp. 89–101. doi: 10.1007/1-4020-4107-1_7.
- Gilmore, R. O. *et al.* (2017) 'Progress toward openness, transparency, and reproducibility in cognitive neuroscience', *Annals of the New York Academy of Sciences*, 1396(1), pp. 5–18. doi: 10.1111/nyas.13325.
- Glaser, B. G. and Strauss, A. L. (2009) *The discovery of grounded theory: strategies for qualitative research*. 4. paperback print.. New Brunswick, NJ ua: Aldine Transaction.
- Grant, R. (2015) 'Identifying HSS research data for preservation: A snapshot of current policy and guidelines', *New Review of Information Networking*, 20(1), pp. 97–103. doi: 10.1080/13614576.2015.1110400.
- Hicks, D. and Holbrook, J. B. (2020) 'A Cartography of Philosophy's Engagement with Society', *Minerva*, 58(1), pp. 25–45. doi: 10.1007/s11024-019-09384-3.
- Higman, R. and Pinfield, S. (2015) 'Research data management and openness', *Program: electronic library and information systems*, 49(4), pp. 364–381. doi: 10.1108/PROG-01-2015-0005.
- Houtkoop, B. L. *et al.* (2018) 'Data Sharing in Psychology: A Survey on Barriers and Preconditions', *Advances in Methods and Practices in Psychological Science*, 1(1), pp. 70–85. doi: 10.1177/2515245917751886.
- Huang, R. (2016) *RQDA: R-based Qualitative Data Analysis. R package version 0.2-8.*, *R-QDA. R-based Qualitative Data Analysis*. Available at: <http://rqda.r-forge.r-project.org/>.
- Kehm, B. M. (2010) 'The Future of the Bologna Process — The Bologna Process of the Future', *European Journal of Education*, 45(4), pp. 529–534. doi: <https://doi.org/10.1111/j.1465-3435.2010.01453.x>.
- Kim, T. (2017) 'Academic mobility, transnational identity capital, and stratification under conditions of academic capitalism', *Higher Education*, 73(6), pp. 981–997. doi: 10.1007/s10734-017-0118-0.
- Knight, G. (2015) 'Building a research data management service for the London school of hygiene & tropical medicine', *Program Electronic Library & Information Systems*, 49(4), pp. 424–439. doi: 10.1108/PROG-01-2015-0011.
- Knorr Cetina, K. (1991) 'Epistemic Cultures: Forms of Reason in Science', *History of Political Economy*, 23(1), pp. 105–122.

- Knorr Cetina, K. (1999) *Epistemic Cultures. How the sciences make knowledge*. 1st edn. Cambridge, Massachusetts: Harvard University Press.
- Kurata, K., Matsubayashi, M. and Mine, S. (2017) 'Identifying the Complex Position of Research Data and Data Sharing Among Researchers in Natural Science', *Sage Open*, 7(3), p. 2158244017717301. doi: 10.1177/2158244017717301.
- Latour, B. (1994) *Science in Action. How to follow scientists and engineers through society*. 6. Aufl. Cambridge, Mass.: Harvard University Press.
- Latour, B. and Woolgar, S. (1986) *Laboratory Life. The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.
- Lefebvre, A., Bakhtiari, B. and Spruit, M. (2020) 'Exploring research data management planning challenges in practice', *it - Information Technology*, 62(1), pp. 29–37. doi: 10.1515/itit-2019-0029.
- Leonelli, S. (2016) *Data-centric Biology. A philosophical study*. Chicago, London: The University of Chicago Press.
- Leonelli, S. (2020) 'Learning from Data Journeys', in Leonelli, S. and Tempini, N. (eds) *Data Journeys in the Sciences*. Cham: Springer International Publishing, pp. 1–24. doi: 10.1007/978-3-030-37177-7_1.
- Leonelli, S., Spichtinger, D. and Prainsack, B. (2015) 'Sticks and carrots: encouraging open science at its source', *Geo*, 2(1), pp. 12–16. doi: 10.1002/geo2.2.
- Lilja, E. (2020) 'Threat of policy alienation: Exploring the implementation of Open Science policy in research practice', *Science and Public Policy*, (scaa044). doi: 10.1093/scipol/scaa044.
- Mancilla, H. A. et al. (2019) 'On a Quest for Cultural Change - Surveying Research Data Management Practices at Delft University of Technology', *Liber Quarterly: The Journal of European Research Libraries*, 29(1), pp. 1–27. doi: 10.18352/lq.10287.
- Mauthner, N. S. and Parry, O. (2013) 'Open Access Digital Data Sharing: Principles, Policies and Practices', *Social Epistemology*, 27(1), pp. 47–67. doi: 10.1080/02691728.2012.760663.
- Mayer, K. et al. (2020) 'Empfehlungen für eine nationale Open Science Strategie in Österreich / Recommendations for a National Open Science Strategy in Austria'. doi: 10.5281/zenodo.4109242.
- Mayernik, M. S. (2016) 'Research data and metadata curation as institutional issues', *Journal of the Association for Information Science and Technology*, 67(4), pp. 973–993. doi: <https://doi.org/10.1002/asi.23425>.
- Mayernik, M. S. (2019) 'Metadata accounts: Achieving data and evidence in scientific research', *Social Studies of Science*, 49(5), pp. 732–757. doi: 10.1177/0306312719863494.
- Mayernik, M. S., Batcheller, A. L. and Borgman, C. L. (2011) 'How institutional factors influence the creation of scientific metadata', in *Proceedings of the 2011 iConference*. New York, NY, USA: Association for Computing Machinery (iConference '11), pp. 417–425. doi: 10.1145/1940761.1940818.
- McKiernan, E. C. et al. (2016) 'How open science helps researchers succeed', *eLife*, 5. doi: 10.7554/eLife.16800.

Mirowski, P. (2018) 'The future(s) of open science', *Social Studies of Science*, 48(2), pp. 171–203. doi: <https://doi.org/10.1177/0306312718772086>.

Münch, R. (2014) *Academic Capitalism*. London, New York: Routledge.

Olesk, A., Kaal, E. and Toom, K. (2019) 'The possibilities of Open Science for knowledge transfer in the science-policy interface', *Journal of Science Communication*, 18(3). doi: 10.22323/2.18030203.

Oushy, M. H. *et al.* (2015) 'To Share or Not to Share? A Survey of Biomedical Researchers in the U.S. Southwest, an Ethnically Diverse Region', *PloS one*, 10(9), p. e0138239.

Pechar, H. and Pellert, A. (2004) 'Austrian Universities Under Pressure From Bologna', *European Journal of Education*, 39(3), pp. 317–330. doi: <https://doi.org/10.1111/j.1465-3435.2004.00186.x>.

Peng, G. *et al.* (2018) 'A Conceptual Enterprise Framework for Managing Scientific Data Stewardship', *Data Science Journal*, 17(0), p. 15. doi: 10.5334/dsj-2018-015.

Piowar, H. A., Day, R. S. and Fridsma, D. B. (2007) 'Sharing Detailed Research Data Is Associated with Increased Citation Rate', *PLOS ONE*, 2(3), p. e308. doi: 10.1371/journal.pone.0000308.

Piowar, H. A. and Vision, T. J. (2013) 'Data reuse and the open data citation advantage', *PeerJ*, 1, p. e175. doi: 10.7717/peerj.175.

Reichmann, S. *et al.* (2020) *Between administration and research - Understanding data management practices in a mid-sized technical university*. preprint. SocArXiv. doi: 10.31235/osf.io/75ac6.

Reuter, J. A., Spacek, D. V. and Snyder, M. P. (2015) 'High-Throughput Sequencing Technologies', *Molecular Cell*, 58(4), pp. 586–597. doi: 10.1016/j.molcel.2015.05.004.

Rhoades, G. *et al.* (2019) 'Marketing to International Students: Presentation of University Self in Geopolitical Space', *Review of Higher Education*, 43(2), pp. 519–551.

Ribes, D. (2017) 'Notes on the Concept of Data Interoperability: Cases from an Ecology of AIDS Research Infrastructures', in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Portland, Oregon, USA: Association for Computing Machinery (CSCW '17), pp. 1514–1526. doi: 10.1145/2998181.2998344.

Ribes, D. (2019) 'STS, Meet Data Science, Once Again', *Science, Technology, & Human Values*, 44(3), pp. 514–539. doi: 10.1177/0162243918798899.

Schulze-Cleven, T. *et al.* (2017) 'The new political economy of higher education: between distributional conflicts and discursive stratification', *Higher Education*, 73(6), pp. 795–812. doi: 10.1007/s10734-017-0114-4.

Shibayama, S. (2012) 'Conflict between entrepreneurship and open science, and the transition of scientific norms', *JOURNAL OF TECHNOLOGY TRANSFER*, 37(4), pp. 508–531. doi: 10.1007/s10961-010-9202-7.

Slaughter, S. and Rhoades (2010) *Academic Capitalism and the New Economy. Markets, State, and Higher Education*. Baltimore, MD: The Johns Hopkins University Press.

Stichweh, R. (1994) *Wissenschaft, Universität, Professionen: soziologische Analysen*. 1. Aufl.. Frankfurt am Main: Suhrkamp (Suhrkamp Taschenbuch Wissenschaft 1146).

Strevens, M. (2006) 'The Role of the Matthew Effect in Science', *Studies in the History and Philosophy of Science*, 37, pp. 159–170. doi: 10.1016/j.shpsa.2005.07.009.

Suber, P. (2012) *Open Access*. MIT Press. Cambridge, MA.

Suchman, L. (2007) *Human-Machine Reconfigurations. Plans and Situated Actions*. 2nd edn. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo: Cambridge University Press.

Tennant, J., Jacques, D. and Collister, L. (2016) 'The academic, economic and societal impacts of Open Access: an evidence-based review', *F1000Research*. Edited by L. Collister, 5. doi: 10.12688/f1000research.8460.1.

Tenopir, C. *et al.* (2011) 'Data Sharing by Scientists: Practices and Perceptions', *PLOS ONE*, 6(6), p. e21101. doi: 10.1371/journal.pone.0021101.

Tenopir, C. *et al.* (2015) 'Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide', *PLOS ONE*, 10(8), p. e0134826. doi: 10.1371/journal.pone.0134826.

Timmermans, S. and Tavory, I. (2012) 'Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis', *Sociological Theory*. doi: 10.1177/0735275112457914.

Toelch, U. and Ostwald, D. (2018) 'Digital open science—Teaching digital tools for reproducible and transparent research', *PLOS Biology*, 16(7), p. e2006022. doi: 10.1371/journal.pbio.2006022.

Unal, Y. *et al.* (2019) 'Research data management and data sharing behaviour of university researchers', *INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL*, 24(1).

Vicente-Saez, R. and Martinez-Fuentes, C. (2018) 'Open Science now: A systematic literature review for an integrated definition', *Journal of Business Research*, 88, pp. 428–436. doi: 10.1016/j.jbusres.2017.12.043.

Wallis, J. C., Rolando, E. and Borgman, C. L. (2013) 'If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology', *PLOS ONE*, 8(7), p. e67332. doi: 10.1371/journal.pone.0067332.

Whitmire, A. L., Boock, M. and Sutton, S. C. (2015) 'Variability in academic research data management practices Implications for data services development from a faculty survey', *PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS*, 49(4), pp. 382–407. doi: 10.1108/PROG-02-2015-0017.

Wilkinson, M. D. *et al.* (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3, p. 160018.

Wilsdon, J. *et al.* (2015) 'The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management'. doi: 10.13140/RG.2.1.4929.1363.