
Article:
Higman, R. and Pinfield, S. (2015) Research data management and openness: The role of data sharing in developing institutional policies and practices. Program, 49 (4). pp. 364-381. ISSN 1758-7301

https://doi.org/10.1108/PROG-01-2015-0005

Research data management and openness: The role of data sharing in developing institutional policies and practices

Rosie Higman and Stephen Pinfield

Abstract

Purpose: To investigate the relationship between research data management (RDM) and data sharing in the formulation of RDM policies and development of practices in higher education institutions (HEIs).

Design/methodology/approach: Two strands of work were undertaken sequentially: firstly, content analysis of 37 RDM policies from UK HEIs; secondly, two detailed case studies of institutions with different approaches to RDM based on semi-structured interviews with staff involved in the development of RDM policy and services. The data are interpreted using insights from Actor Network Theory.

Findings: RDM policy formation and service development has created a complex set of networks within and beyond institutions involving different professional groups with widely varying priorities shaping activities. Data sharing is considered an important activity in the policies and services of HEIs studied, but its prominence can in most cases be attributed to the positions adopted by large research funders.

Research limitations/implications: The case studies, as research based on qualitative data, cannot be assumed to be universally applicable but do illustrate a variety of issues and challenges experienced more generally, particularly in the UK.

Practical implications: The research may help to inform development of policy and practice in RDM in HEIs and funder organisations.

Originality/value: This paper makes an early contribution to the RDM literature on the specific topic of the relationship between RDM policy and services, and openness – a topic which to date has received limited attention.

Keywords: research data management, research data services, open data, data sharing, openness, actor network theory

1. Introduction

Researchers in Higher Education (HE) are producing ever-increasing quantities of digital data in the course of their work which needs to be managed for both immediate and potential long-term use (Borgman, 2012; Pryor, 2012). This data, the production of which has been facilitated by technological changes which have transformed the research process over the last 20 years, is being used and sometimes reused in innovative new ways (Hine, 2006). Reuse of data is, however, still by no means routine in many disciplines, despite its potential (Borgman, 2012; Lynch, 2008); and this has led some to call for data sharing to be written into the policies of funders and institutions (Royal Society, 2012). Many research funders, which in the UK include the publicly-funded Research Councils and also charities (such as the Wellcome Trust), now require their grant holders not only to adhere to certain standards of data management but also commit wherever possible to data sharing (Research Councils UK, 2011).

Given the emphasis research funders are placing on sharing, it is surprising that there is little discussion in the literature relating to Research Data Management (RDM) of the relationship between data sharing on the one hand, and developing RDM policies and practices at institutional level on the other hand. In particular, little attention has been given to the tension between research funders' emphasis on data sharing and the apparent reluctance of many researchers to share their data (Borgman, 2012; Wallis et al., 2013) and how this is being played out in institutions. Examining these tensions and relationships is crucial to understanding how and why data sharing is contributing to the development of RDM policies and emerging practices.

This research aims to examine what is driving the formation of RDM policies and practices in UK Higher Education Institutions (HEIs), and the role played by data sharing in this process. This aim is addressed, firstly, by analysing institutional RDM policies across HEIs and, secondly, focusing on case studies of two particular institutions. The analysis is reported using Actor Network Theory (ANT) as a lens through which to view the phenomenon of RDM, helping to make sense of the complex relationships at play. An explanation of ANT is provided in the literature review which precedes the policy analysis and case studies, and which also provides a survey of current trends in RDM in general and data sharing in particular.

2. Literature Review

2.1. The scope and importance of data and RDM

In a commonly-cited and wide-ranging definition, "data" is characterised as "facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors" (National Research Council, 1999, p. 15 in Borgman, 2012). Data is commonly gathered or produced as part of the academic research process and is now being generated in ever-increasing volumes and in a wide variety of often quickly-superseded digital formats (Berman and Cerf, 2013; Borgman, 2012; Pryor, 2012). This creates a need for active data management before datasets deteriorate (Cox and Pinfield, 2013; Higgins, 2012; Lavoie, 2012). Complementing this practical need is a change in perceptions of the value of research data: it has come to be viewed as an asset which should be managed to sustain its value (Borgman,

2012; Carlson and Garritano, 2010; Lavoie, 2012). There is an awareness, for example, of the potential benefits of being able to submit data as part of national research evaluation exercises (Cox and Pinfield, 2013). In addition, more negative concerns around risk avoidance can also drive RDM developments, including those around researchers complying with Freedom of Information (FOI) legislation (e.g. H.M. Government, 2000) and the potential costs of them not doing so (Whyte and Tedds, 2011). All of these factors create incentives at an institutional level for HEIs to engage with RDM.

Furthermore, research funders emphasize how managing and sharing datasets can make more cost-effective use of (often publicly-funded) research grants (Lavoie, 2012; Pryor, 2012). Reflecting this, many funders' research data policies effectively place the onus for RDM on institutions rather than individual researchers to ensure effective RDM takes place (Brown and White, 2014; Jones, 2012). In the UK, the policy of the largest Research Council, the Engineering and Physical Sciences Research Council (EPSRC, 2011), in particular has "galvanized institutions" but, as Pryor (2014a) observes, there is a danger that the funder requirements will be seen as merely an "administrative hurdle" (Pryor, 2014a, p. 22), unless wider arguments about the importance of RDM for good research are made and cultural change is achieved.

2.2. Why share data?

The emphasis research funders are placing on sharing data, as well as the need to manage data effectively, aligns with wider trends in higher education towards openness, most prominent in the movement towards Open Access (OA) to research publications (Finch et al., 2012; Suber, 2012; Willinsky, 2006). Data sharing has been promoted for many reasons: firstly, without sharing data it is impossible to verify the results of research, a key principle of good science (Borgman, 2012). For others it is a political issue: withholding data generated with public funds is seen as undemocratic and it would be wrong to restrict access to a public good (Arzberger et al., 2004; Murray-Rust, 2008; Vision, 2010). A less altruistic argument is made that data sharing can increase a researcher's citation rate, whether by direct citations of the data or of the associated article (Brase, 2014; Pampel and Dallmeier-Tiessen, 2014; Piwowar and Vision, 2013).

Despite these varied arguments in favour of data sharing, there are relatively low levels of sharing currently occurring (Pryor, 2014b). There are also marked disciplinary differences in the level of sharing, partially caused by the diverse types of materials used in different disciplines and significantly, given the costs involved in preparing data for public release (Borgman, 2012), the availability of funding in different disciplinary areas. More generally, there are few direct incentives for individual researchers (Pampel and Dallmeier-Tiessen, 2014), especially when research funder requirements are not yet being strictly enforced (Jones, 2012), or they are early in their careers and trying to establish reputations (Hine, 2006). There are also concerns, for both researchers and HEIs, that the costs of RDM and data sharing are poorly understood and funders are offering only minimal financial support (Vision, 2010).

Despite these drawbacks, research funders' policies mean that there is a growing awareness of the importance of RDM in many HEIs, with, for example, libraries reporting institutional RDM policy

development being a high priority in most UK universities (Cox and Pinfield, 2013). These policies respond to a number of drivers, including storage and security, but also data sharing (Pinfield, Cox, et al., 2014). Policies are variously used to provide "credentials" for those championing RDM (Pryor, 2014a), gain access to funding for IT infrastructure (Jones, 2014), clarify institutional positions (Brown and White, 2014), and outline roles and responsibilities (Brown and White, 2014). However, the policies are not without difficulties. There are, for example, concerns that they will become unachievable statements of aspiration as opposed to intent (Jones, 2014). Even when policies are written realistically there remains a significant risk that they will be unfulfilled without both considerable cultural change (Brown and White, 2014; Pryor, 2014a) and acceleration of the implementation of promised services (Pryor, 2014b).

2.3. Theoretical background

Given the complex array of incentives and disincentives for the different actors involved in RDM policies and practices, it is helpful to establish a theoretical perspective from which to view and analyse the relationships involved. Actor Network Theory (ANT) has the potential to play this role, focusing as it does on relationships within different contexts in order to explain organisations and structures by tracing the links between different actors (Latour, 2005). It has been applied in a wide variety of contexts, including the study of OA and institutional repositories by Kennan and Cecez-Kecmanovic (Kennan and Cecez-Kecmanovic, 2007; Kennan, 2011). They deploy the idea of "programs" and "anti-programs" in HEIs to explain the conflicting forces acting on those involved in the implementation of an OA service, and how the balance of these forces influences the success of the initiatives in institutions (Kennan and Cecez-Kecmanovic, 2007; Kennan, 2011). An approach like this could ostensibly be applied to RDM efforts at many HEIs, with research funders as the main external actors, as opposed to the publishers in Kennan's analysis (Kennan, 2011). ANT offers potential insights into both OA and RDM due to its focus on tracing connections beyond the usual boundaries of a group to explain the composition of networks of actors in the context of new innovations (Latour, 2005), and how these connections are framed to create apparently coherent, contained organisations and markets (Callon, 1999). By emphasising the social sphere as the tracing of associations rather than as a specific context (Latour, 2005), ANT becomes a powerful explanatory tool for RDM which, like OA, involves professional groups extending beyond their usual relationships to develop services in response to a large-scale challenge (Lewis, 2010).

ANT theorists seek to redefine the "social" as, "nothing other than patterned networks of homogeneous materials" (Law, 1992, p. 381). Therefore, 'actors' within such a social network can be humans, animals, technologies, or indeed documents, all "intertwined in constantly changing relationships" (Silvis and Alexander, 2014, p. 111). As Silvis and Alexander (2014) observe, this position is "a logical outcome of the rejection of all of the following: techno-deterministic views; beliefs that the human actors fully control non-human actors; and assumptions that non-human actors on neutral." (Silvis and Alexander, 2014, p. 111). It therefore offers interesting insights for information science researchers. The range of possible actors in ANT is further expanded by the idea of "punctualised" networks; common patterns normally simplified as single actors, such as a funding agency, but which are impermanent so may be disaggregated into its constituent parts in times of change thus becoming multiple actors (Law, 1992).

This growth in the range of potential actors, including groups and objects, has significant implications for the analysis of institutions, such as universities. Policy documents, for example, cease to be the end products of the interaction between a series of human actors, and become independent actors in their own right capable of changing human behaviour.

ANT attempts to refocus on the interactions between actors rather than researchers' assumptions of what the actor does (Latour, 1999, p. 19). In this context-dependent approach actors only take on meaning as part of a 'network' and their significance shifts as networks change (Latour and Woolgar, 1986, p. 107). The term network is employed in order to escape the ideas of proximity and boundaries (Latour, 1996), and its limitless nature means an analysis is likely to involve tracing only a limited part of that network. Both Latour (1996) and Law (1999), early proponents of ANT, stress the heterogeneity of relationships, returning to the importance of recognising complexity and not sacrificing an accurate description for the sake of simplicity (Law, 1999).

Emphasising the changing nature of relationships, Law (1999) argues the relationships between actors are actively performed and thus what is important to understand is how relationships, these performances, become durable over time. By understanding the determinants of a durable relationship it may be possible to predict better how a practice such as RDM, where there are many new and possibly fleeting connections between actors, is likely to evolve and which relationships will endure to shape new activities. In order to understand what sustains a performed relationship, it is necessary to trace links from initially being established in a new scenario, through a period of stabilisation (frequently through documentation), to being a composed assemblage in a wider context (Latour, 2005). The analysis below attempts to begin to do this for RDM in HEIs.

2.4. Research questions

The existing literature, therefore, identifies why RDM is important, suggests reasons why and how researchers share their data, and discusses potential benefits (and limitations) of RDM policies. However, little work has been done on the details of what is actually included in institutional RDM policies and why decisions to include particular issues were made, a gap in the literature this article aims to begin to address specifically for the UK. This leads to the primary research question addressed in this study:

- To what extent does data sharing influence the formation of RDM policies and emerging practices in UK Higher Education Institutions?

This gives rise to supplementary questions:

- What are the drivers for the development of RDM policies and emerging practices in UK HEIs?
- To what extent are universities writing RDM policies primarily to fulfil funders' requirements?
- To what extent is openness an enabler or constraint when it comes to researchers engaging with managing their data?
- How much does the linking of RDM and openness vary among UK HEIs?

3. Methodology

Two forms of data collection and analysis were used sequentially in this study to answer the research questions. Firstly, quantitative and qualitative analysis of RDM policies available from UK HEIs was undertaken. Secondly, qualitative analysis of interviews conducted at two case study institutions was carried out. The research approach adopted an interpretivist perspective, assuming that, "social reality exists as part of human experience and is socially constructed" (Cecez-Kecmanovic and Kennan, 2013). This assumption aligns well with ANT, where meaning is dependent on the surrounding network. By positing reality as being dependent on context, as opposed to a single objective reality, it reduces the importance of reconciling conflicting viewpoints, but instead focuses upon understanding different perspectives and their origins.

In total, 37 UK RDM policies were analysed (listed in Appendix 1), gathered from the Digital Curation Centre (DCC) website and by searching for 'research data management policy' and 'research data management policy for universities' on Google.co.uk, considering the first 100 results, accurate up to 19 July 2014. This represents a comprehensive survey of the UK Higher Education (HE) sector and was made more focused by excluding roadmaps and strategies rather than policies. As this policy analysis was intended to provide an early indication of key trends to inform the rest of the study, the analysis focused initially on basic quantitative techniques as well as qualitative topical analysis (Bryman, 2012; Richards, 2009). This was followed by a more inductive process of coding the policies alongside the case study interview transcripts.

The two case study institutions were chosen for their contrasting approaches to RDM policy formation. Analysing cases with distinct approaches makes it possible to clarify what is influencing decisions, how different actors are behaving and how networks are being formed. Case 1 was an exemplifying case (Bryman, 2012), a medium-sized research-intensive institution with a recently finalized RDM policy, ostensibly forming part of the "late majority" in (Rogers, 1962, 2003) typology of the diffusion of innovations. Case 2 was an extreme case (Bryman, 2012), an "innovator" (Rogers, 1962, 2003), a large research-intensive university which was among the first UK HEIs to produce an RDM policy and has led national projects developing exemplary RDM services and infrastructure. At both institutions, staff from all three key professional services involved in delivering RDM services (librarians, research office staff and IT professionals) were interviewed, with a mixture of senior staff, concerned with strategic direction of the overall service, and those who are delivering the RDM service day to day. In total, 11 interviews were conducted (6 in Case 1 and 5 in Case 2).

All but one of the interviews were conducted in person (the remaining one, by telephone) between 19 June and 15 August 2014. Interviews took a semi-structured approach to allow the issues pertinent to each institution, profession, and individual to be thoroughly explored – with interviews averaging 37 minutes. The research approach was approved by the University of Sheffield Information School Research Ethics Panel and, in line with this, written, informed consent was given by all participants and

the data they provided was anonymized during transcription. This data was coded using the NVivo software; initially creating numerous, frequently topical, codes (Richards, 2009), before developing more analytical codes focusing on how respondents linked themes (Braun and Clarke, 2006), and bearing in mind the ANT perspective.

In spite of the systematic procedures described above there are several methodological limitations to the research. Perhaps most significantly, it needs to be recognised that the results of the case studies cannot necessarily be seen as representative of the UK HE sector as a whole (Thomas, 2011). They are inevitably the reflections of the personal views of specific individuals in particular institutions, albeit ones directly involved with RDM. Instead of necessarily representing the entire sector, the experience of the two institutions may, however, be confidently said to show some of the varying incentives and disincentives with which institutions are generally faced and provide examples of how they are responding to the challenges. Furthermore, the research approach adopted was designed to highlight a wide range of issues by including two contrasting case studies and combining them with a cross-sector analysis of policies, in an attempt to triangulate results.

4. Results

4.1. Policy analysis

RDM policies of UK HEIs, definable in ANT as distinct actors, are representations of the most common concerns and the attitudes towards RDM taken in institutions. The majority of the policies analysed in this study were published in 2012 or 2013, and are typically linked to other policies within that institution (most commonly concerning research ethics, information security and intellectual property rights). This situates RDM within a broader research governance agenda, a network of policies, research administrators, senior managers and academics. Many policies lacked detail and specificity, with only 23 of the 37 having a named owner, 20 stating the aim of the policy, and 14 actually defining what they meant by 'research data'. Furthermore only 8 policies even vaguely addressed how RDM was to be funded, and even fewer explicitly acknowledged the aspirational nature of what was being outlined – a critical issue given most services are still in their infancy. Whilst a lack of detail is understandable at this early stage in the development of policies it does potentially limit their influence within nascent RDM networks.

RDM policies did, however, highlight the importance of research funders, suggesting RDM networks which frequently span institutional boundaries. RCUK were the most commonly referenced external institution appearing in just over half (20) of the policies analysed, including several specific references to the RCUK Common Principles (Research Councils UK, 2011), an influential actor in the space. This compares with under a third (12) citing the Digital Curation Centre and only 4 citing another external body. Data sharing was the most commonly cited driver of RDM (see Table 1), but in many cases it was explicitly linked to funder requirements, indicating that for the authors of the policies this activity was only significant due to other actors in the RDM networks framed around their institutions.

Research Data Management Drivers	Number of Policies	Percentage
Source of Funding for RDM Activities	8	22%
Intellectual Property Rights	20	54%
Security	31	84%
Funders' Requirements	33	89%
Data Sharing	36	97%

Table 1: The number and proportion of research data management policy documents mentioning key drivers for RDM (n = 37)

Data management planning was the most frequently mentioned stage in the research data lifecycle (see Table 2). Since DMPs are commonly required by research funders (Jones, 2012), this suggests again research funders' requirements are critical in the networks of those writing RDM policies. This is apparently corroborated by looking at the frequent mentions of preservation in nearly all of the RDM policies considered, usually framed in terms of funders' requirements. Furthermore, positive statements in relation to data sharing were often accompanied by caveats highlighting intellectual property rights, commercial interests, and ethical issues, with a tone that sometimes discouraged sharing beyond meeting funders' requirements:

> *"In the event that research data created at UCA is **required** to be released for regulatory and/or contractual requirements this will be in accordance with appropriate safeguards." (The University for the Creative Arts (2012) Research Data Management Policy) [emphasis added]*

This can be contrasted with the University of East London (2012) where there is a "presumption of releasing data", followed by the usual range of caveats.

Stages of the Research Data Lifecycle	Number of Policies	Percentage
DMPs	37	100%
Active Data Management	17	46%
Disposal	14	38%
Preservation	36	97%

Table 2: The number and proportion of policy documents which mentioned key stages of the research data lifecycle (n = 37)

The other key life-cycle stages mentioned related to data preservation and disposal. Several policies mentioned keeping data for a set period since the last citation, or, more generally, if the dataset was judged to be significant. Similarly, disposal was usually discussed in the context of legal and ethical issues suggesting concerns about protecting institutional reputations as well as previously mentioned drivers.

> *"Research Data shall be retained for longer than 10 years…where the results of the research become contentious or subject to challenge at any time during the initial 10 year retention period, in which case Research Data should be retained pending review and not…disposed of until the matter is fully resolved." (University of Southampton, 2012)*

4.2. Case studies

Creating an RDM policy was amongst the first key developments reported at both case-study institutions but their approaches to this task differed considerably. In Case 1, the policy was produced largely by one individual in the Research Office, with only brief consultations with other professionals and academics. In contrast, in Case 2 there was an 18-month period of consultation with internal stakeholders, led by the library, before the legal services team finalized the wording. This consultation was prompted by internal drivers for RDM: research governance, preservation and security, which were sufficient to encourage action prior to the EPSRC announcement of its policy. It was believed that a policy would have been developed in the institution regardless of funder requirements:

> *"I think we can honestly say we wanted to develop one for ourselves, because when we started …it was just pre-EPSRC discussions so a lot of this was about…other issues. Things like Climategate at UEA... So that was what was hot on people's agenda, have we got enough governance of this area?" (Case 2 Senior Library Manager)*

In contrast, in Case 1 there was awareness of the need for RDM but the policy would not have been in place without the "stick" 'threatened' by the EPSRC:

> *"It's the threat of the punishment that does tend to activate and concentrate the mind" (Case 1 Research Office Director)*

Where funders were influential in Case 2 was in the prioritisation of the development of services to researchers, where much activity revolved around supporting DMPs (Case 2 Library Staff and IT Staff).

Open Data was barely mentioned as a driver of RDM policies in both cases but was frequently discussed with regards to implementing services. Interviewees at both institutions linked data sharing to OA to published outputs, asserting that attitudes to the latter were likely to shape their response to the former. This was seen as beneficial in Case 2, which was an early adopter of OA, with researchers thought to be "very used to flying…the Open Access flag" (Case 2 Senior Library Manager), and this

meant there was less resistance to Open Data. However, in Case 1, the library was still advocating OA to researchers, and awareness of, and support for, data sharing, was also believed to be low.

> *"…we haven't gone in there talking about Open Data as such, because I think for many people it is a freakish thought" (Case 1 Library Staff)*

These differences were likely to be exaggerated at the time of interview when Case 1 was still establishing researchers' levels of knowledge via a survey and interviews, whilst there was already an RDM enquiry service, DMP support, and RDM training, amongst other activities in Case 2. However, there were differences in how staff at the two institutions understood the relationship between Open Data and RDM. In Case 1, they were described as clearly separate issues, while two interviewees in Case 2 described open data sharing as the end point of a continuum of sharing activities which begins with sharing with oneself i.e. RDM.

Despite the differences between the cases there was agreement that many researchers' objections to data sharing often originate in misunderstandings about, for example, embargo periods and exceptions available for commercial or ethical issues.

> *"So it's trying to convince them that we are not…trying to steal, manipulate, in some way take ownership of their work" (Case 1 IT Professional)*

In Case 2, where advocacy had taken place, it was assumed that at least some objections had been addressed. However, support staff interviewed were generally in contact with researchers who ask for assistance: an atypical group with high levels of RDM awareness who are likely to be more receptive than the average researcher to RDM developments. This fact made it hard to verify whether objections are reduced after a period of education.

RDM policies were seen at both institutions as a statement of their current position and what should be done. There were concerns that RDM is currently an unfunded mandate in Case 2 and that it would be problematic from a resources perspective if the policy was adopted systematically:

> *"If every single researcher abided by the policy, all the dots and crossed t's,…we would have a massive problem, we would not have the capacity to do it." (Case 2 Library Staff)*

In contrast, in Case 1, there were some frustrations from those not directly involved in producing the policy that it lacked ambition, and in particular that previous work on preservation had been ignored in favour of a policy which solely aimed to fulfil funders' requirements (Case 1 Senior Library Manager).

Policies were then in place at both sites, and the institutions were working towards compliance with funders' requirements, but the role of policy within the institutions was not well defined. It was clear that data sharing was important as an issue at both HEIs but the reasons for its prominence and the

relationship between the concepts of RDM on the one hand and openness on the other differed significantly.

5. Discussion

5.1. Overview

Evidence presented in this study from both strands of data collection suggests that data sharing is influential in the formation and development of RDM policies and practices, but that in many HEIs this issue is primarily seen in terms of research funder requirements. Other drivers of RDM (which are strictly speaking in ANT terms themselves particular sorts of actors) include preservation and research governance, and they are also encouraging institutions to address the challenge of rapidly-rising quantities of research data, but are still conceived in abstract ways and lack obvious practical implications. The perceived importance of these other drivers varies between institutions, but they are generally insufficient in themselves to prompt action on RDM. In contrast, funders' requirements appear to be encouraging greater engagement with RDM, and, to some extent, are aligning the incentives of administrators, support staff and researchers. Research funder policies are becoming important actors (in ANT terms) in their own right. They have a direct impact on the views of researchers and support services staff on RDM and help to shape institutional activities which indirectly may influence success in research grants applications. Formal institutional RDM policies further cement this relationship between funders' and funding recipients, making these temporary relations more durable and creating in turn a new actor, the institutional policy, which is likely to influence the implementation of RDM in the future.

5.2. Data sharing

The fact that data sharing has often been perceived in terms of compliance with funder policies is indicated by the frequent coincidence of the issue in relation to funder requirements in institutional policies. Different attitudes are, however, evident in the case studies. Sharing seems to be driving some practices in Case 2, whilst discussions about Open Data are being delayed in Case 1 until RDM is more accepted. Some of these differences may be a result of how Open Data is understood: if it is an 'add-on', rather than integral, to RDM, it is much easier to demote. In Case 2, data sharing seems to be part of an RDM punctualised network; that is, data sharing, security and preservation among other issues are commonly aggregated and presented as a single actor, RDM. Whereas in Case 1, data sharing and RDM are situated separately in the network, although their link to each other is likely to be strengthened once RDM has been integrated into researchers' workflows. This suggests that where RDM and Open Data are viewed as separate issues, external funders are more significant in determining the adoption of data sharing.

OA and RDM and Open Data are frequently supported by many of the same staff within HEIs and whilst the University of Exeter (2013) is exceptional in linking these issues at a policy level, there seems to be a clear link for those implementing RDM. Hence the decision in Case 1 to take a gradual approach to RDM as there is still some resistance to OA, whilst Case 2 can be more proactive due to their previous successes at promoting OA. The similar networks which can be traced for OA and RDM support mean

that success in the former may enable the implementation of the latter, shaping the ways an institution approaches RDM.

The attitudes of researchers to data sharing, however, inevitably vary. Therefore, it is helpful to consider the incentives and resources available to researchers in explaining their behaviours as actors. There can be considerable costs associated with preparing data for release so, whatever their views in principle, many researchers require a direct incentive to share their data, and as citation of datasets is not yet widespread, research funder requirements are currently the main incentive to share.

5.3. Research funders, institutions and their policies

The interviewees outlined a wide range of drivers responsible for the growth of RDM at UK HEIs, but the EPSRC (EPSRC, 2011) statement in particular greatly increased RDM's perceived strategic importance for senior managers. In moving responsibility from individuals to institutions, EPSRC raised the stakes associated with a single researcher mismanaging their data (Brown and White, 2014). Simultaneously, there is evidence in both strands of the research of growing concern within institutions about the governance of research data prompted by the freedom of information (H.M. Government, 2000) and data protection (H.M. Government, 1998) legislation, contributing to the rise of RDM higher up the agenda of UK HEIs.

The production of DMPs was seen as being significantly influenced by research funders, even though DMPs were also viewed as useful in their own right (in terms of planning at an individual and institutional level). All interviewees in Case 2 remarked upon the large increase in DMPs produced by the medical faculty coinciding with several funders in this area requiring grant holders to produce a DMP. This demonstrates the power of funding bodies, as actors separate from their policies, to change behaviours through enforcement. The effective implementation of RDM requirements is vital to avoid "inviting lip service to be paid where implementation is undesirable or impractical" for researchers (Jones, 2012), as they have sometimes been with OA (Pinfield, Salter, et al., 2014). However, compliance monitoring (and consequently, enforcement) is perhaps easiest in areas such as production of DMPs, whilst other areas are more difficult to monitor and therefore less likely, in the short term at least, to be enforced. Therefore, the data sharing requirements of funders may not easily spread beyond disciplines where sharing is already common, due to the absence of incentives for researchers, high costs for institutions and lack of sanctions for non-compliance.

Whilst research funders' RDM policies feature in discussions between support staff and researchers, institutional policies it appears are often marginalized, a situation exacerbated by many being vague and aspirational. Institutional RDM policies seem to largely be advocacy documents used at the centre of institutions, produced to lobby senior stakeholders, gain access to resources (Pryor, 2014a), and also satisfy external funders. However, once they have been created these policies do become separate actors, influencing RDM decisions in their own right. Sometimes, however, especially in view of their lack of clarity, these policies can create a new set of governance challenges. For example, a lack of clarity

regarding archiving data in the policy may actually create confusion about where researchers should deposit data for preservation, especially given the rapid service development occurring in this area.

5.4. Other drivers

Whilst recognizing the influence of research funders and their policies, there are other drivers for RDM related to technological challenges, legal concerns and governance issues. At a practical level the rapid proliferation of research datasets has created a sizeable storage problem for researchers and HEIs, which is often only vaguely addressed in RDM policies. The need to move researchers away from self-managed storage to using centrally-managed storage facilities, is causing concern about how to provide and fund sufficient resilient, networked storage to departments with wildly differing requirements and budgets. Although this issue is barely mentioned in policies it is a key one for researchers and is likely to become more prominent as institutions move into RDM service delivery.

In contrast, long-term storage, often addressed at length in policies, is not a current priority in practice at either case study site. Due to the large costs involved with preserving data there needs to be a clear process for deciding what data is of sufficient value to be retained, curated and preserved (Lavoie, 2012). Library staff in Case 2 identified the phrase 'significant' data as the key to their policy being sustainable and the beginning of their selection process, although this clearly requires further definition. The impetus for this activity, outside of funders' requirements, seems to largely originate in library services, but unless it can be effectively linked to research governance processes it may be difficult to generate the necessary resources for effective data preservation.

Good research governance involving oversight of policies and processes, of which research data governance is a part, allows HEIs to both demonstrate the high quality of their research (Case 1 Research Office Director) and manage the risk of reputational damage from mismanaged data (Hickman, 2012), making it an important driver of RDM. The main concerns expressed in both the interviews and the policies analysed (complying with the FOI Act (H.M. Government, 2000), Data Protection Act (H.M. Government, 1998), and ethical issues) all highlight the need for effective RDM throughout the research process, regardless of whether or not researchers intend to share their data. The possibility of reputational damage and legal action resulting from poor research governance can elevate the perceived importance of RDM, especially with senior management, in terms of the institutional priorities of UK universities.

5.5. ANT models

The networks traced at the two case study institutions reflect the different approaches taken to RDM and openness as well as the different stages they have reached in developing their services. At both institutions, actors such as funders, academic competition within their field, and potential commercial uses of their data, among others, are combining in complex ways to produce a network of heterogeneous links between researchers (across different disciplines and career stages), support services and policy documents.

Figures 1 and 2 attempt to capture the key relationships for the two case study institutions and illustrate differences between them, using a simple ANT "graphical syntax" (Silvis and Alexander, 2014) in which actors (the library, IT services etc) are mostly depicted within boxes, although (to make them clear) policy documents (institutional RDM policy and research funder policies) are pictured with 'document graphics'. RDM and openness, the focus of this study, are depicted as overlapping but distinct core agendas (in themselves actors) with various other actors relating to them. The ways in which actors relate to the core agendas are indicated by arrows with the 'weight' of the arrows indicating the strength of the relationship between the actors and the core agendas. Programs (solid arrows) indicate an enabling relationship whilst anti-programs (dashed arrows) are obstructing relationships (as in Kennan, 2011). This is, of course, a simplified view which aims to foreground the relationships between the different actors (particularly organisational units and policy documents) specifically in relation to RDM and openness. General relationships between these different actors are not depicted and, similarly, relationships within the different actors (each of which is in itself a complex network) are assumed to exist but not depicted. Future studies could usefully unpack some of these complexities.

Perhaps the most noticeable point illustrated by the Figures is the stronger RDM programs relating to the actors in Case 2 (Figure 2) compared with Case 1 (Figure 1). In Case 2, the policy was created after considerable consultation and discussion, and took into account views of different actors. In contrast, in Case 1, there is less clear and consistent support for RDM, and there was only limited input into the policy documents from various actors. Most importantly for this study, the Figures illustrate the different approaches to RDM and openness. In Case 1, the relationship between RDM and openness is itself a weak one compared with Case 2. There are also clearer anti-programs evident in Case 1 in relation to openness from a number of actors, particularly researchers, and whilst this is also evident in Case 2, it is less prominent. The RDM and openness agendas are seen in Case 2 to be more naturally aligned.

Where the networks and approaches to RDM and data sharing differ, this seems to relate partly to who is leading the development of policies and services. In Case 1, the Research Office led the policy development, and the policy and planned activities reflect their concerns about meeting funder requirements. In Case 2, the initial impetus came from academics with support from senior managers, and the library has since coordinated the RDM activity with a broader focus on research governance. These patterns are consistent with data reported by Pinfield, Cox and Smith (2014). The process of policy development has also created different networks in Case 2 where there was a highly consultative approach to the policy, as recommended by Jones (2014), leading to academic champions in a range of departments who are key links promoting RDM to their colleagues.

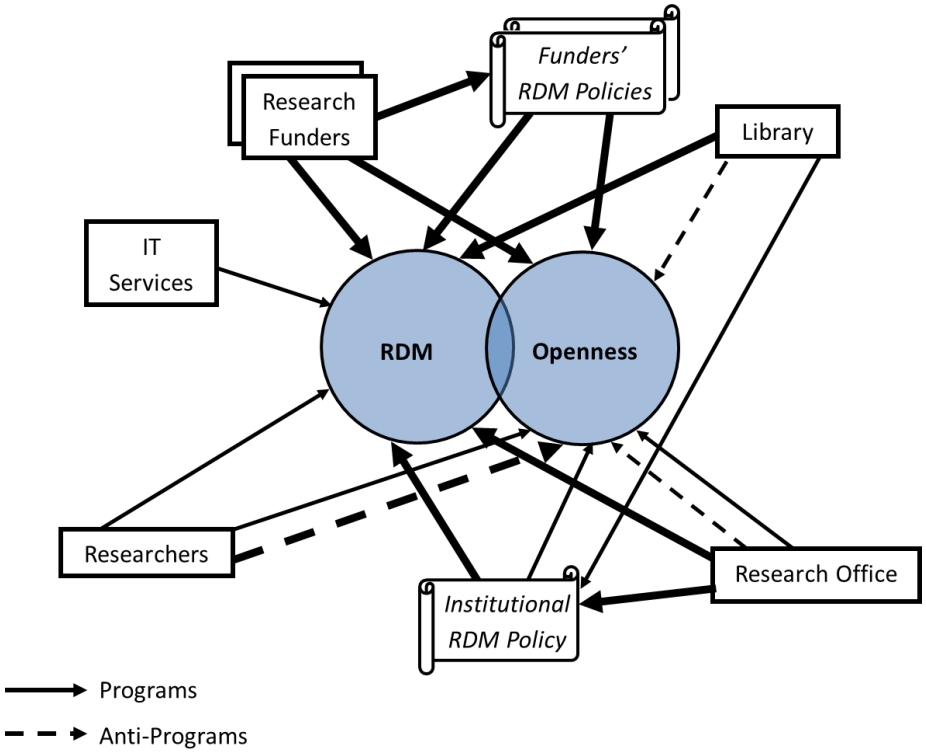

Figure 1: Case 1: Programs and anti-programs relating to RDM and openness by actor

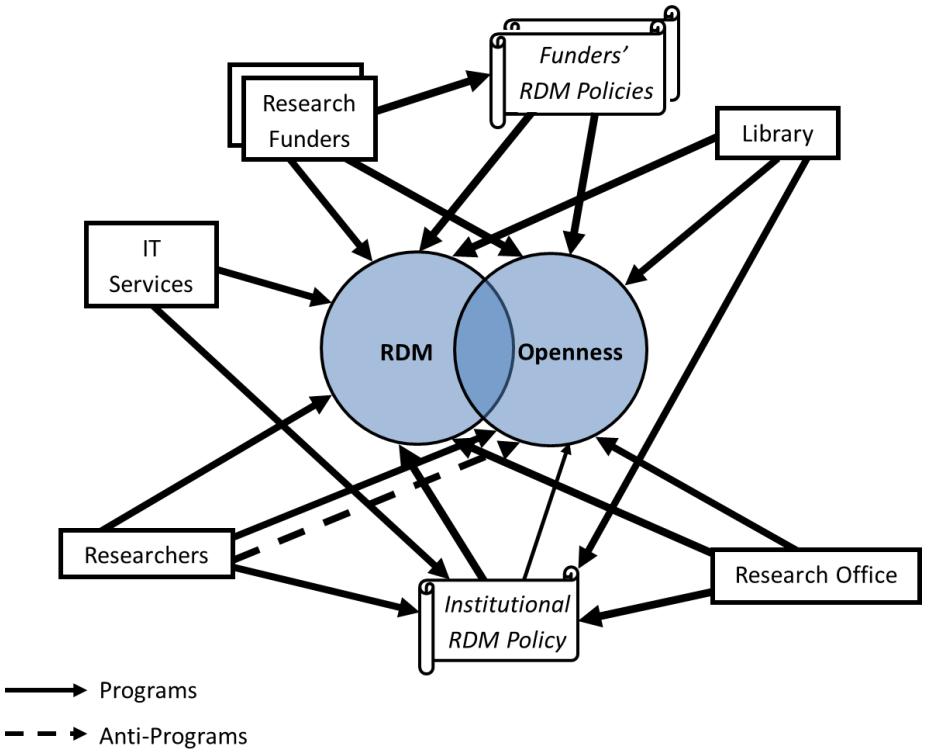

Figure 2: Case 2: Programs and anti-programs relating to RDM and openness by actor

6. Conclusion

Tracing the networks currently operating across the two case study institutions reveals the range of incentives influencing the implementation of RDM, and how these vary according to previous activities, the profession leading efforts, and institutional priorities. Data sharing is an important factor in the majority of the policies analysed and in the activities at both HEIs studied. However, for the majority of institutions, its prominence can be attributed to the requirements of large research funders. The exception to this situation is found at institutions, such as Case 2, where there is a strong recent history of adopting openness, initially in relation to published outputs and particularly in certain faculties, and so data sharing is a more fundamental part of current RDM practice.

However, aside from funders' requirements, which are not currently being enforced fully and so are having less influence on researchers' practice than on institutional decision making, there are currently few definite incentives for researchers in most disciplines to share data. A possible area for future research would be the attitudes of researchers at institutions where a prolonged period of advocacy promoting Open Data has already occurred, such as in Case 2, to see if there was a discernible impact on researchers' attitudes. There have been many assessments of researchers' views at the beginning of the development of RDM services at different institutions but few after researchers have been trained and supported in managing and possibly sharing their data. This may reveal whether researchers have fundamental objections to sharing data or whether many merely require reassurance and support, as suggested by several interviewees in this research.

The idea of documents as actors, advanced by ANT, seems to find support in the empirical evidence presented here of the influence funder policies are having on approaches to RDM. Even though they are not currently fully enforced, these policies are significantly influencing the content of institutional policies, and priorities of support service staff and researchers, either directly or indirectly. Policies are the most easily accessed and interpreted statement of funders' position on RDM, and are clearly being considered by HEIs when devising their own policies. Although this study suggests institutional policies are not currently as influential as those of research funders, they do act as defined and accessible statements of institutions' positions and are likely to continue to be important points of reference as RDM developments continue. The extent to which they develop to encourage greater openness in relation to data is an important point which remains to be seen and is, at least for the foreseeable future, likely to vary across different institutions with different networks of actors.

Acknowledgements

Thanks for Dr Mary Anne Kennan for her very helpful comments on a draft of this article.

References

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., et al. (2004), "Promoting Access to Public Research Data for Scientific, Economic, and Social Development", *Data Science Journal*, Vol. 3 No. November, pp. 135–152.

Berman, F. and Cerf, V. (2013), "Who will pay for public access?", *Science*, Vol. 341 No. 6146, pp. 616–7.

Borgman, C.L. (2012), "The conundrum of sharing research data", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 6, pp. 1059–1078.

Brase, J. (2014), "Making data citeable: DataCite", in Bartling, S. and Friesike, S. (Eds.),*Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*, Springer International Publishing, Cham, Vol. 445, pp. 327–329.

Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, Routledge, Vol. 3 No. 2, pp. 77–101.

Brown, M.L. and White, W. (2014), "Case Study 2: University of Southampton - a partnership approach to research data management", in Pryor, G., Jones, S. and Whyte, A. (Eds.),*Delivering Research Data Management Services*, Facet Publishing, London, pp. 135–162.

Bryman, A. (2012), *Social research methods*, Oxford University Press, Oxford, 4thed.

Callon, M. (1999), "Actor-network theory—the market test", *The Sociological Review*, Vol. 47 No. 1, pp. 181–195.

Carlson, J. and Garritano, J. (2010), "E-science, cyberinfrastructure and the changing face of scholarship: Organizing for new models of research support at the Purdue University libraries", in Walter, S. and Williams, K. (Eds.),*The Expert Library: Staffing, Sustaining, and Advancing the Academic Library in the 21st Century*, Association of College and Research Libraries, Chicago, IL, pp. 234–269.

Cecez-Kecmanovic, D. and Kennan, M.A. (2013), "The methodological landscape: Information systems and knowledge management", in Williamson, K. and Johanson, G. (Eds.),*Research Methods: Information, Systems, and Contexts*, Tilde Publishing, Prahran, Victoria, pp. 113–138.

Cox, A.M. and Pinfield, S. (2013), "Research data management and libraries: Current activities and future priorities", *Journal of Librarianship and Infiormation Science*, Vol. 0 No. 0, pp. 1–18.

EPSRC. (2011), *EPSRC policy framework on research data*, Engineering and Physical Science Research Council, Swindon, available at: http://www.epsrc.ac.uk/about/standards/researchdata/ (accessed 1 May 2014).

Finch, J., Bell, S., Bellingan, L., Campbell, R., Donnelly, P., Gardner, R., Hall, M., et al. (2012), *Accessibility, sustainability, excellence: How to expand access to research publications.*, Research Information Network, London, available at: http://www.researchinfonet.org/wp-content/uploads/2013/02/Final-version.pdf.

H.M. Government. (1998), "Data Protection Act 1998."

H.M. Government. (2000), "Freedom of Information Act 2000."

Hickman, L. (2012,April), "University of East Anglia spent £112,870 on 'climategate' PR", *The Guardian*, London, available at: http://www.theguardian.com/environment/blog/2012/apr/24/uea-climate-change-email-publicity.

Higgins, S. (2012), "The lifecycle of data management", in Pryor, G. (Ed.),*Managing Research Data*, Facet Publishing, London, pp. 17–46.

Hine, C. (2006), "Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work", *Social Studies of Science*, Vol. 36 No. 2, pp. 269–298.

Jones, S. (2012), "Research data policies: Principles, requiremets and trends", in Pryor, G. (Ed.),*Managing Research Data*, Facet Publishing, London, pp. 4–66.

Jones, S. (2014), "The range and components of RDM infrastructure and services", in Pryor, G., Jones, S. and Whyte, A. (Eds.),*Delivering Research Data Management Services*, Facet Publishing, London, pp. 89–114.

Kennan, M. and Cecez-Kecmanovic, D. (2007), "Reassembling scholarly publishing: Institutional repositories, open access, and the process of change", *Australian Conference on Information Systems: The 3 Rs: Research, Relevance and Rigour - Coming of Age, December 5-7, 2007*, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1030964 (accessed 15 April 2015).

Kennan, M.A. (2011), "Learning to share: Mandates and open access", *Library Management*, Emerald Group Publishing Limited, Vol. 32 No. 4/5, pp. 302–318.

Latour, B. (1996), "On actor-network theory: A few clarifications", *Soziale Welt*, Vol. 47 No. 4, pp. 369–381.

Latour, B. (1999), "On recalling ANT", in Law, J. and Hassard, J. (Eds.),*Actor Network Theory and after*, Blackwell Publishers, Oxford, pp. 15–25.

Latour, B. (2005), *Reassembling the social: An introduction to Actor-Network Theory*, Oxford University Press, Oxford.

Latour, B. and Woolgar, S. (1986), *Laboratory life: The construction on scientific facts*, Princeton University Press, Princeton, New Jersey.

Lavoie, B.F. (2012), “Sustainable research data”, in Pryor, G. (Ed.),*Managing Research Data*, Facet Publishing, London, pp. 67–82.

Law, J. (1992), “Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity”, *Systems Practice*, Vol. 5 No. 4, pp. 379–393.

Law, J. (1999), “After ANT: Complexity, naming and topology”, in Law, J. and Hassard, J. (Eds.),*Actor Network Theory and after*, Blackwell Publishers, Oxford, pp. 1–14.

Lewis, M. (2010), “Libraries and the management of research data”, in McKnight, S. (Ed.),*Envisioning future academic library servcies: Initiatives, ideas and challenges*, Facet Publishing, London, pp. 145–168.

Lynch, C. (2008), “Big data: How do your data grow?”, *Nature*, Nature Publishing Group, Vol. 455 No. 7209, pp. 28–29.

Murray-Rust, P. (2008), “Open data in science”, *Serials Review*, Vol. 34 No. 1, pp. 52–64.

National Research Council. (1999), *A question of balance: Private rights and the public interest in scientific and technical databases*, National Academic Press, Washington DC.

Pampel, H. and Dallmeier-Tiessen, S. (2014), “Open research data: From vision to practice”, in Bartling, S. and Friesike, S. (Eds.),*Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*, Springer, Cham, pp. 213–224.

Pinfield, S., Cox, A.M. and Smith, J. (2014), “Research data management and libraries: Relationships, activities, drivers and influences.”, *PLoS ONE*, Public Library of Science, Vol. 9 No. 12, p. e114734.

Pinfield, S., Salter, J., Bath, P.A., Hubbard, B., Millington, P., Anders, J.H.S. and Hussain, A. (2014), “Open-access repositories worldwide, 2005-2012: Past growth, current characteristics, and future possibilities”, *Journal of the Association for Information Science and Technology*, Vol. 65 No. 12, pp. 2404–2421.

Piwowar, H.A. and Vision, T.J. (2013), “Data reuse and the open data citation advantage”, *PeerJ*, Vol. 1, p. e175.

Pryor, G. (2012), “Why manage research data?”, in Pryor, G. (Ed.),*Managing research data*, Facet Publishing, London, pp. 1–16.

Pryor, G. (2014a), “Options and approaches to RDM service provision”, in Pryor, G., Jones, S. and Whyte, A. (Eds.),*Delivering Research Data Management Services*, Facet Publishing, London, pp. 21–40.

Pryor, G. (2014b), “A patchwork of change”, in Pryor, G., Jones, S. and Whyte, A. (Eds.),*Delivering research data management services*, Facet Publishing, London, pp. 1–19.

Research Councils UK. (2011), “RCUK common principles on data policy.”

Richards, L. (2009), *Handling qualitative data: A practical guide*, Sage, London.

Rogers, E.M. (1962), *Diffusion of innovations*, The Free Press, New York.

Rogers, E.M. (2003), *Diffusion of innovations*, Simon & Schuster International, New York, 5thed.

Royal Society. (2012), *Science as an open enterprise: Final report*, Royal Society, London, available at: https://royalsociety.org/policy/projects/science-public-enterprise/Report/ (accessed 1 May 2014).

Silvis, E. and Alexander, P.M. (2014), "A study using a graphical syntax for actor-network theory", *Information Technology and People*, Vol. 27 No. 2, pp. 110–128.

Suber, P. (2012), *Open access*, MIT Press, Boston, MA, available at: http://mitpress.mit.edu/books/open-access.

Thomas, G. (2011), *How to do your case study: A guide for students and researchers*, Sage Publications Ltd, London.

University for the Creative Arts. (2012), *Research data management policy v7*.

University of East London. (2012), *Research data management policy for UEL*.

University of Exeter. (2013), "Open access research and research data management policy", University of Exeter.

University of Southampton. (2012), "Section IV: Research data management policy."

Vision, T.J. (2010), "Open data and the social contract of scientific publishing", *Bioscience*, Vol. 60 No. 5, pp. 330–331.

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.", (Nunes Amaral, L.A.,Ed.)*PloS one*, Public Library of Science, Vol. 8 No. 7, p. e67332.

Whyte, A. and Tedds, J. (2011), *Making the case for research data management*, Digital Curation Centre, Edinburgh, available at: http://www.dcc.ac.uk/webfm_send/487.

Willinsky, J. (2006), *The access principle: The case for open access to research and scholarship*, MIT Press, Cambridge, MA.