The 2014 DAF Survey at the University of Sheffield

Andrew Cox
The Information School
University of Sheffield

Laurian Williamson
Library Service
University of Sheffield

Abstract

The Data Asset Framework methodology has evolved to provide a model for institutional surveys of researchers' data practices and attitudes. At least 13 such studies have been published in the UK and internationally. The aim of this paper is to analyse the results from the 2014 DAF survey at the University of Sheffield and to reflect on the comparability of this with previous published studies. 432 researchers responded to the survey representing 8% of the target population. Researchers at Sheffield collect multiple types of data and a significant number have accumulated very large amounts of data. Data was backed up on a diverse basis. Only 25% of respondents had a DMP. Eighteen months after its creation most respondents were still not aware of the local research data management policy. Fortunately, most respondents were favourable to the idea of training in many aspects of RDM. Researchers had generally had no experience of sharing data, but attitudes were positive, both in terms of a significant minority seeing a lack of data sharing as an obstacle to the progress of research and also desire to reuse the data of others and share their own with a broad group of researchers. Comparison of the Sheffield results with those of other institutions is difficult particularly because of the divergence of questions asked in the different studies. Nevertheless, in terms of data practices and identifying training priorities there are common patterns. This institutional survey showed less positive attitudes to data sharing than the results of cross-institutional studies, such as conducted by Tenopir et al. (2011).

Introduction

One of the challenges of Research Data Management (RDM) is our lack of a detailed understanding of existing data practices. The sheer variety of types of data created, ways of using this data and data sharing practices across different disciplinary and sub-disciplinary communities make understanding institutional support requirements complex. In the process of developing Research Data Services (RDS) a key part of discovering requirements is to gather institutional level data (Whyte, 2014). The Data Asset Framework (DAF) is one method that has been developed to provide framework for collecting data (Jones et al., 2008; Data Asset Framework, 2009; HATII, 2009). This is a flexible methodology for creating case studies and gathering more general survey information about local research data practices. Indeed, it is increasingly referenced as the main source for institution wide surveys of researchers; results from more than a dozen such surveys have been published openly 2010-2014, and it is likely many others been conducted but not been published. The primary purpose of such surveys is to inform local RDS development and support service requirements, and to provide evidence for targeted advocacy and outreach. Conducting a survey raises the RDM agenda within the institution. Questions often directly inform respondents, e.g. alerting them to the existing RDM policy. Contact details of RDM staff supplied on the survey may generate first contact with researchers who feel they have an immediate problem. Each survey may also help to build up cross-institutional data to identify common patterns and for benchmarking. By offering a standard approach, DAF lowers the threshold to investigating local practices, potentially validates survey questions and it also creates scope for comparison and aggregation of data across institutions to build up a bigger picture. Benchmarking local experience against relevant peers would seem to be a reasonable aspiration for such surveys. Understanding how far the local practices and attitudes are typical would contribute to making policy decisions. In this context the purpose of this paper is to report the 2014 Sheffield DAF survey of researcher data practices, awareness and attitudes, and to investigate the comparison of the results to patterns found in other published studies.

The Data Asset Framework

Central to developing effective RDS is its fit to what researchers need for their own purposes and may require in order to fulfil the requirements from their funder or publisher. Discipline-specific studies paint part of the picture; cross organisational, cross-disciplinary surveys of researchers have also been conducted, e.g. Tenopir et al. (2011). The data profiling methodology provides a systematic framework for capturing comprehensive descriptions of existing research data sets, akin to the case study aspect of DAF (Witt et al., 2009). However, the development of RDS implies investigating data practices and attitudes at the institutional level.

The Humanities Advanced Technology and Information Institute (HATII) led an EU project that developed and tested a methodology for conducting case studies of research data use, originally called the Data Audit Framework, then the Data Asset framework[1].

---

[1] Data Asset framework: http://www.data-audit.eu/index.html

This was a systematic but practical approach to capturing information about data collections that had been created in the institution.

> 'The Data Asset Framework is a set of methods to: find out what data assets are being created and held within institutions; explore how those data are stored, managed, shared and reused; identify any risks e.g. misuse, data loss or irretrievability; learn about researchers' attitudes towards data creation and sharing; suggest ways to improve ongoing data management.' (HATII, 2009)

Increasingly, however, studies referencing the DAF approach have been primarily online questionnaire-based surveys of all research-active staff and other key stakeholders in data management, presumably partly because such surveys are vastly less resource intensive to conduct and partly because they supply a level of data across the institution, rather than focusing on a few cases. Implicit in the DAF approach is the idea of adapting the investigation to the needs of the institution.

Table 1. DAF studies published 2010-2014.

Study	When main study conducted	Responses (response rate, where known)		Report supported by interviews	No. questions
Northampton	June/July 2010	80		16	32
Georgia Tech (USA)	2010-13 (sic)	77		26	23
Southampton	May 2011	239		No	30
Exeter	Feb 2012	284		50	34
Lincoln	Apr 2012	44	(8%)	No	23
Hertfordshire	May-July 2012	67	(12%)	No	24
Nottingham	July-Sep 2012	366		No	20
London School of Hygiene and Tropical Medicine	Aug-Sep 2012	117	(16%)	No	15
Leeds	July-Nov 2012	242		No	18
Oxford	Nov 2012	314		No	21
UEL	Nov 12-Jan 2013	51		No	34
Humboldt (Germany)	Jan-Mar 2013	499	(24%)	No	24
Essex	Feb 2013	55		No	31
Sheffield	Jan 2014	433	(8%)	No	23

Table 1 provides summary details of the published DAF studies published up until 2014 (Alexogiannopoulos et al., 2010; Grace, 2013; Knight, 2013; Nassiri and Worthington, 2012; Open Exeter Project Team, 2012; Parsons et al., 2013; RoaDMaP, 2013; Rolando et al., 2013; University of Hertfordshire, 2012; University of Lincoln, 2012; University of Southampton, 2011; Van den Eynden, 2013a, 2013b; Wilson, 2013; Wilson et al., 2012). There have been other institutional studies, for example, the

University of Newcastle's Iridium project conducted a similar type of survey but by project rather than individual researcher; several US institutions have published surveys based on a different set of questions (e.g. Akers and Doty, 2013). However, DAF has emerged as the standard reference point for such investigations, at least in the UK.

Research Questions

This paper reports the results of a survey of University of Sheffield researchers in 2014, broadly organised around six research questions:

1. What are the data practices of researchers in terms of types and quantity of data collected, where data is stored and how it is backed up?

2. What awareness do researchers have of the RDM policy?

3. What training needs do researchers perceive they have?

4. What is their experience and attitude to data sharing?

In addition, the authors explored two other questions:

5. How do the Sheffield results compare with that of other institutions?

6. How comparable have been the published DAF surveys?

Methodology

Sheffield is a research intensive institution; a member of the Russell Group of Universities. The Director of Library Services and University Librarian, Martin Lewis, has been ahead of his time in advocating the importance of RDM and has played an important role in building University support for RDM (Lewis, 2010). In 2012 the University Research Data Management Policy was approved and embedded within the new University Good Research and Innovation Practices (GRIP) policy. As a result of an internal RDM project in 2011/2012, a full-time Research Data Management Coordinator position was created. At the time of conducting the survey, in January and February 2014, the university was committed to developing RDM support services and tools but was still at an early stage of culture change and pushing RDM up agenda as a strategic priority. There was a support web site, a help desk email, but as yet no institutional data repository.

The RDM survey was conducted by a newly established University Research Data Management Service Delivery Group, consisting of key professional service stakeholders and an academic representative. This was the first RDM survey at the University of Sheffield exploring issues around a research data management infrastructure (technical and human) to support the research data lifecycle, acknowledging and responding to different practices across subject disciplines. In order for the Research Data Management Service Delivery Group to scope and plan for a sustainable institutional RDM service it was necessary to understand what kind of research data was held by researchers, and to identify the influences and barriers to managing research data, where support and training were required, and the current levels of RDM practice in the faculties. The information provided in the survey was

used to inform the ongoing institutional RDM user requirements gathering exercises, provide an initial assessment of what research data we were seeking to manage, and help identify what tools, infrastructure, and policies were required in order to facilitate and embed good RDM practice at The University of Sheffield.

The selection of questions used in the survey was close to those used for the Nottingham study. Additionally, with permission, several questions used by Tenopir et al. (2011) about data sharing attitudes were added. The complete list of questions are available in Appendix 1.

The survey was online hosted on a locally installed instance of LimeSurvey and open to all research active staff. It was promoted through a number of channels, such as through the general announcement list, engaged key stakeholders including the library, research office, and IT support staff, and via the Faculty Directors of Research. The analysis is based on the 432 complete responses.

After exploring the figures it was concluded that the data for many questions was best analysed by Faculty and staff role. Responses were categorised by tenured academic (professor, assistant professor, reader, senior lecturer, lecturer) (N=164), PhD students (N=126), contract researcher (research associate and research assistant) (N=74) and other, which included research fellows and data managers (N=66). This excludes one PhD that could not be located in a faculty and one academic from an international faculty.

Table 2. Response rates per faculty.

Faculty	Tenured academics			Research associates and RAs			PhD students		
	Total	Responded	Rate	Total	Rsp	Rate	Total	Rsp	Rate
Faculty of Arts and Humanities (ARTS)	176	10	6%	19	4	21%	322	25	8%
Faculty of Engineering (ENG)	282	23	8%	223	14	6%	978	38	4%
Faculty of Medicine, Dentistry and Health (MED)	255	59	23%	246	42	17%	395	17	4%
Faculty of Science (SCI)	251	31	12%	239	12	5%	648	13	2%
Faculty of Social Sciences (SOCSCI)	341	41	12%	71	2	3%	664	33	5%
	1305	164	13%	798	74	9%	3007	126	4%

35% of all responses were from the Faculty of Medicine Dentistry and Health (MED) (N=153); and a third of these were from one school. MED responses were fairly well distributed across roles. Responses from other faculties were lower and less evenly spread, e.g. 50% of Faculty of Arts and Humanities (ARTS) respondents were PhD students. This seemed primarily to reflect a better response rate, though there are some

structural differences between faculties e.g. the ratio of staff to PhD students in Engineering is approaching 1:2; whereas in MED the ratio is 2:1. Research associates are bigger groups in Faculty of Science (SCI) and Faculty of Engineering (ENG), and to a lesser degree MED, than in the other two faculties. Over 50% of contract researcher (research associate and research assistant) responses were from MED. A high proportion of Faculty of Social Sciences (SOCSCI) respondents were from one department (21%). Given that the faculty has 13 different departments it is probably the least well represented in the data.

The total response rate was 8% (430/5451), though the figures for total numbers in the institution have to be treated with a little caution, e.g. a number of figures existed for registered PhDs in the university. The total 5451 used to calculate the response rate excludes teaching staff, since they were not invited to complete the survey. Interestingly, response rates were better from tenured staff compared to others, with the lowest response rates from PhDs. It needs to be considered whether respondents represent the wider population in terms of attitudes; on the expectation that those more interested in a topic will reply to a survey, and those who are more interested will be better informed and perhaps have better practices. If this were the case it would imply that the topic is relatively uninteresting to PhD students, although the low overall response rate from this group might be more related to issues with communications when promoting the survey to them, rather than how they viewed the topic. Follow up work should clarify this.

Looking at where staff (N=299) were gaining funding for their current research projects (PhD students were excluded from analysis) the top funders were: The UK research funding councils (38% of all 533 mentions), charities (18%), EU (9%) and commercial organisations (7%). Many respondents mentioned the University itself as a funder, but this might refer to personal research effectively funded by employment by the university rather than formally funded research, as only ten people responded indicating that their research had no funder and only another ten identified their research as self-funded. An interesting finding of the survey is the wide range of funders. Any expectation that most researchers are funded by RCUK (and so would be heavily influenced by their RDM policy) would be unfounded.

Data Practices

Question 8 looked at the formats of data researchers were creating (Figure 1). Inevitably, this is a broad brushstroke question that covers formats (such as documents) and some broad categories of data (e.g. transcripts). For one set of data a respondent could tick multiple boxes. For example, an interview study could easily generate audio files, documents, transcripts and potentially spreadsheets.

Nearly everyone was creating certain basic formats of document: 91% of people created documents; 76% spreadsheets. There were some distinct patterns across other types of data: for example 44% of people in ARTS used film or photographs; it was less than 25% in other faculties. 76% of engineers used models, algorithms and scripts; only 20% of all other respondents used them. 19% of respondents collected artefacts, specimens etc. This was far less common in SOCSCI. On average people collected five of the forms of data mentioned in the questionnaire, 42% of respondents had over six forms of data – reinforcing the point that most researchers collect multiple formats of data, which is one reason why RDM is complex.
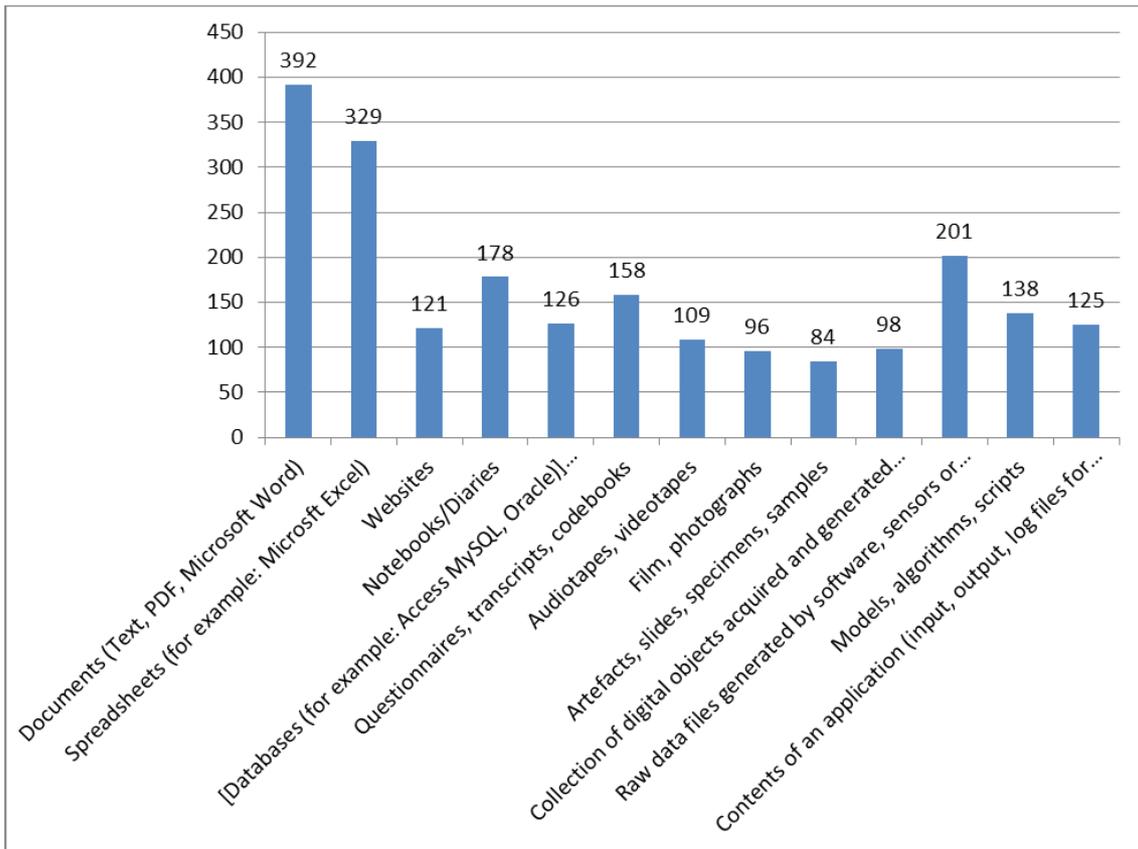
Figure 1. Responses to the question about types of data researchers create or work with.
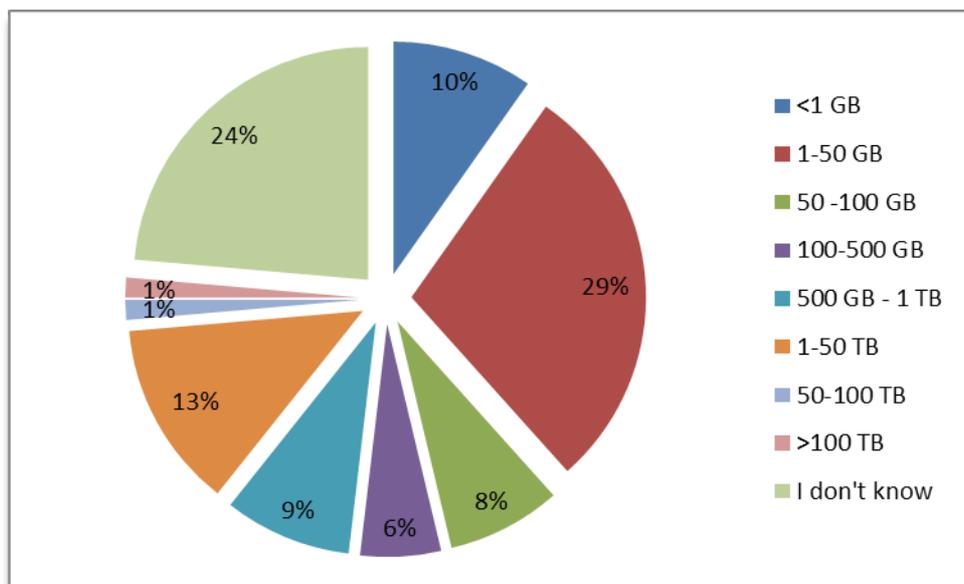

Figure 2. Researchers' estimates of the volume of research data they were expecting to create during their current research activities.

Another question asked how much data researchers were expecting to create (Figure 2). 39% estimated 0-50 GB. This implies that for this large group there may be issues around the fragmentation of data storage, but they have not yet entered the realms of "big data" in the sense of sheer volume (though given the wording of the question they could have large amounts of retrospective data). However, 15% said they expected to create more than a terabyte. Of these, 29 were in engineering (32% of respondents from that faculty), 16 from science (27%) and 17 from MED (11%). This seems to confirm the emergence of big data volumes at least in science. 24% responded that they did not know the answer to the question. Given that researchers have been involved in multiple projects, it is perhaps surprising that three quarters of respondents did feel confident to give an estimate.

Respondents were also asked where they stored data "during the life" of their project. Top answers were: portable storage device (65%), local disk (61%), and web based services (44%). Surprisingly, the latter came above university drive, although this is slightly unclear because Sheffield offers two forms of data storage: a Novell-based drive and a Google drive. Answers for these two storage places were: University drive 30%; University Google drive 21%. Only 9% said they were storing their data in a departmental run data management system. 35% said they kept data on paper: a salutary reminder of the importance of material data storage (and this was apparent across respondents from all faculties). On average, people named four different places where they were storing live data. This could be because they have different types of data stored in different places, but does suggest quite complex approaches to storage, with resilience through multiple backups, but risks in terms of version control. There was some evidence of variability across faculties: for example, the MED used the university drive more than other faculties.

28% said that they were backing up daily; 17% weekly. 38% backed up on an ad hoc basis – presumably backing up data when it was created. Only 3% said never; perhaps these people were relying on backing up by central systems. 5% did not know. Perhaps only the 5% who did not know would be an immediate cause for concern. Divergence of data back-up practices could easily reflect divergent patterns of data collection. Unlike routine backing up of work, data collection is often intermittent, so an ad hoc strategy could be appropriate.

A further question asked respondents whether they had a Data Management Plan. Although researchers might have had different practices in different projects, few took the opportunity of the "other" option to suggest this. 25% of respondents said they had a DMP; 65% said they did not. 4% said they did not know. The proportion who did have DMPs was a little higher in MED and SOCSCI across all roles. More SCI academics had a DMP (42%), compared to 13% of all PhD students. Rates among PhD students in ARTS, ENG and SCI were low. The relatively low rate of formal DMP is not very surprising.

Overall, 57% respondents said there were legal, regulatory or confidentiality issues shaping how they managed their data. Respondents in MED, and to a slightly lesser degree SOCSCI, were more likely to answer yes.

Policy Awareness

The University of Sheffield published an RDM policy in July 2012, as part of its Good Research and Innovation Practice guidelines[2]. Eighteen months later, only 33% of respondents said they were aware of the policy prior to reading the survey. If nothing else the survey drew respondents' attention to the existence of a policy. Awareness was especially low in Engineering and Science. A higher proportion from MED and SOCSCI said they had heard of the local policy (43% and 44% respectively). Tenured academics had no more awareness than PhD students, except in ARTS where they were much more aware than PhD students. In the Faculty of Science, PhD students were actually more aware than tenured academics.

In responding to a question about policy awareness, 38% of all respondents said they were aware of their funders' requirements; 47% said they were not aware. Figures were lower for ARTS and also among all PhD students, perhaps because they did not have direct funders. Academics in MED, SCI and SOCSCI were much more aware of funder requirements.

Respondents were asked to comment on how long they were required to keep data after the end of the project (Figure 3).
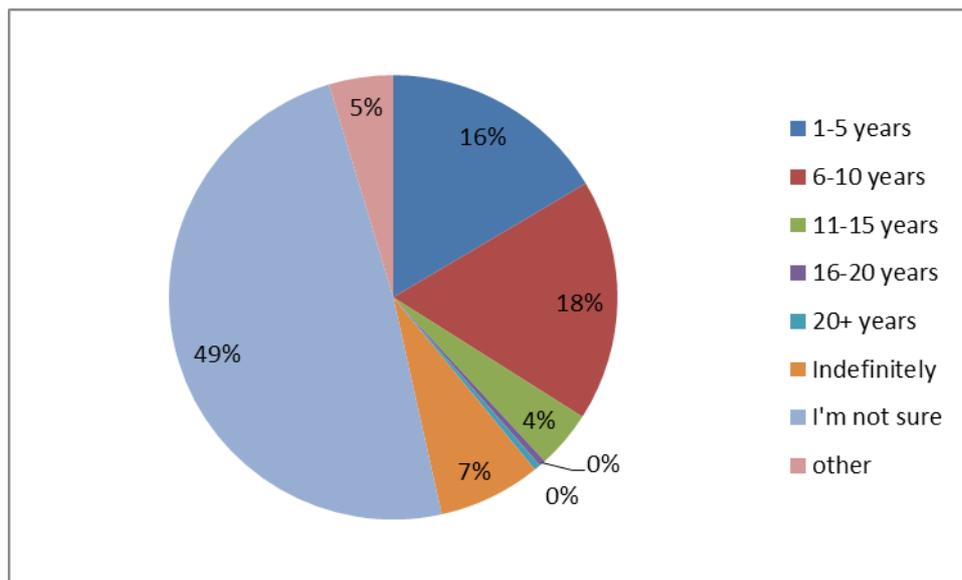

Figure 3. Data retention requirements.

Because researchers work on multiple projects there might be different rules for different projects, but few took the "other" option to explain this. The stand out result is that 49% of respondents said they were not sure about how long they were required to retain data. This points to a significant area where researchers are unclear about the rules under which they have to operate.

---

2 Good Research and Innovation Practice guidelines: http://www.shef.ac.uk/ris/other/gov-ethics/grippolicy/practices/all/rdmpolicy

Training

A majority of respondents were definitely or possibly interested in training on all the topics listed in the questionnaire. This suggests a broad demand for training, though meeting this need is challenging in terms of who should deliver this and how to deliver it across disciplines. Engineering academics were consistently less interested than others in training; engineering PhD students tended to be more in line with average responses. PhD students as a whole were consistently more interested in training than established academics.

The responses to this question could also be taken to suggest major areas of concern for researchers themselves. Thus storage, data management plans, copyright and documentation seem to be areas of strong concern and to a broad group of researchers. Training about funders' requirements is quite low on the list, but is on a par with other topics in terms of the number seeing it as a concern. Ethics could be low on the list because researchers already felt knew enough about this. Metadata was of relatively less interest, though there was a strong interest in documentation, perhaps only because it is a less understood concept.

Table 3. Training needs, ranked by numbers "definitely interested".

Training subject area	Might be interested	Definitely interested	Total with any interest
Storing your research data	36%	36%	72%
Developing a research data management plan	44%	30%	74%
Copyright and Intellectual Property	40%	30%	70%
Documenting your research	43%	29%	72%
Citing your research data	38%	28%	66%
Sharing your research data	46%	25%	71%
Funders requirements and RDM	46%	21%	68% (after rounding)
Creating metadata for research data	36%	21%	57%
Ethics and consent	35%	19%	54%

Data Sharing

Responses to a question about experience of data sharing (Figure 4) suggested that the majority of researchers had not deposited their research data in a dedicated subject or disciplinary repository, though responses varied a little by faculty. 38% of SCI respondents had deposited data, mostly because they were required to. The figure was even higher for academic staff in that faculty, with 52% of them having done so. Indeed, overall 30% of tenured academics had; whereas only 10% of all students had. This could be partly because academics were more likely to work in a team; whereas PhDs usually worked independently and are only likely to deposit data at the end of the project. Depositing data in a repository is only one form of data sharing, so these results

are not necessarily inconsistent with the generally positive view of data sharing apparent in answers to other questions. Yet, given that most funders now mandate data deposit, this figure points to a large gap.
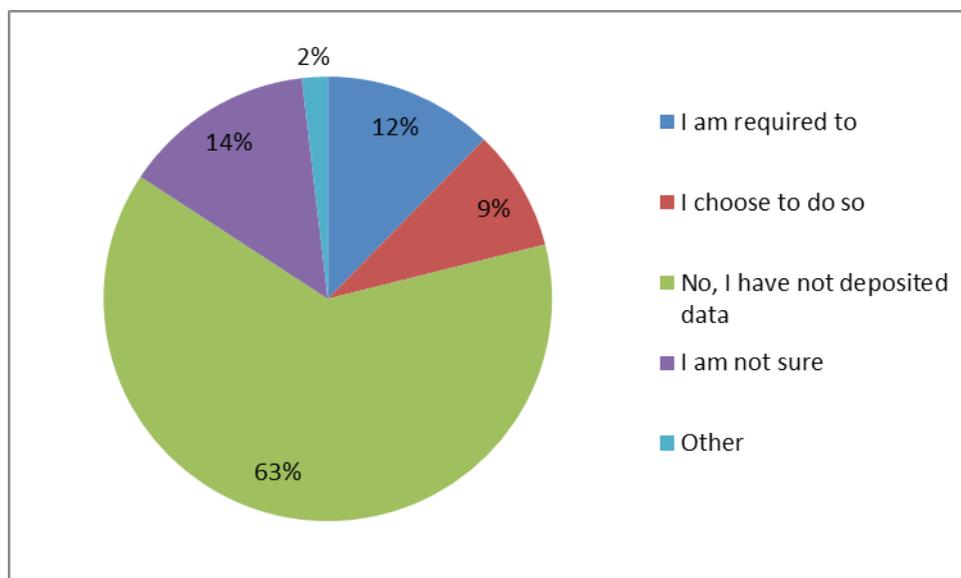

Figure 4. Responses to the question: "Have you, or your research team, ever deposited your research data in a dedicated subject/disciplinary repository?"

In response to the next question, 29% of respondents said they planned to make their data "publicly discoverable and accessible" at the end of the project. This is a little higher than the numbers saying they had done it in the past in the previous question. 33% said they did not plan to do so and 26% said they did not know. The latter is possibly the significant figure when compared to answers from the previous question, signalling uncertainty about depositing data, even though they had not done so before. The response was fairly even across faculty and role: though staff in SCI and PhD students in MED seemed to have more plans to share data.

43% of respondents agreed (strongly or somewhat) with the proposition that: "Lack of access to data generated by other researchers or institutions is a major impediment to progress in my subject discipline." Although less than half of respondents, it does suggest that a significant proportion of researchers see greater openness as a good thing; even if 33% disagreed with the statement. 50% of all PhD students agreed with the statement. Faculty in ARTS and ENG were relatively more sceptical.

In contrast, only 31% of respondents agreed (strongly or somewhat) with the proposition that: "Lack of access to data generated by other researchers or institutions has restricted my ability to answer research questions." With only 7% agreeing strongly with this question, the sense that lack of data sharing blocks individuals' research is quite limited. 40% of respondents positively disagreed with the statement. However, there were marked differences across faculty. ARTS respondents were most strongly in agreement across roles. PhD students were more in agreement, except in SOCSCI. ENG academics were the most sceptical. The prevalence of these sort of attitudes must be seen as representing a significant barrier to any RDM initiative.

The last question in the survey asked respondents to give their level of agreement to a number of statements relating to data reuse. There was not much evidence of differences in response to these questions across roles or faculties; in itself quite

interesting, especially given some variations in responses to questions about data sharing experience and plans.

64% of respondents said they would use other researchers' datasets if they were easily accessible, though 14% actively said they would not do so. Presumably they would not consider that the data of others' would be relevant. 56% of respondents were willing to share their data in a central data repository (the question was ambiguous about whether this was national or institutional); but it was clear that for most respondents there needed to be restrictions on its use – although 17% of all respondents said they would be willing to deposit all their data without restrictions. A similar proportion (62%) were willing to share data "across a broad group of researchers". Thus the numbers wishing to reuse the data of others and willing to share their data were around the same. Nearly two thirds of respondents agreed strongly that reused data should be cited, pointing to the need for better understanding of how to cite data.

Table 4. Attitudes to data sharing.

Statement	Agree strongly	Agree somewhat	All those agreeing	All those actively disagreeing	Tenopir et al. (2011): Strongly agree
I would use other researchers' datasets if their datasets were easily accessible	23%	41%	64%	14%	43%
I would be willing to place at least some of my data into a central data repository with no restrictions	19%	37%	56%	22%	42%
I would be willing to place all of my data into a central data repository with no restrictions	4%	17%	21%	59%	15%
I would be willing to share data across a broad group of researchers	19%	43%	62%	13%	37%
It is important that my data is cited when used by other researchers	64%	23%	87%	4%	69%

Discussion

The picture of RDM practices among researchers at Sheffield is one of great variation and only the early stages of an organisational culture attuned to RDM issues. Researchers at Sheffield collect multiple types of data and a significant number have accumulated very large amounts of data. Live/active data seemed to be stored in multiple places, perhaps for different types of data, but perhaps also as part of a back up strategy. A third of respondents had live data stored in paper form. Data was backed up

on a diverse basis. Only a quarter of respondents had a DMP. Eighteen months after its creation most respondents were still not aware of the local RDM policy. Nearly half of respondents were also unaware of funder requirements or data retention requirements. Fortunately, most respondents were favourable to the idea of training in many aspects of RDM. The strength and breadth of response could be taken to point to major areas of concern being with data storage, DMPs and IP. Researchers generally had no experience of sharing data, but attitudes were positive, both in terms of a significant minority seeing a lack of data sharing as an obstacle to the progress of science and also reporting a desire to reuse the data of others and share their own with a broad group of researchers.

The main benefit of doing such a survey was that the research data management co-ordinator could use the survey results as a tool for engaging a wide range of research staff and students. An example of such engagement was the delivery of individual and tailored presentations to Faculty Directors of Research and Innovation, highlighting areas where there was already good practice within their discipline and faculty, and identifying areas where more subject-specific training and advocacy was required. This engagement of senior research staff is essential when considering the long-term sustainability and resourcing of institutional RDM services and tools. The survey results were also used by key professional service stakeholders from the IT services in their business case to secure additional resourcing and investment in research data management.

Concrete figures and data visualisations of the questionnaire data were found to have a dramatic impact both with senior research managers and the managers of other professional services. This applies particularly to questions about practical aspects of RDM, such as the regularity and processes of backing up data, where active research data was being stored, the volume of data created during the research process, and the lack of awareness of funder expectations regarding retention periods for research data and the existence of the high level institutional research data management policy. Managers were struck by the variability among staff and some evidence of latent demand for data services that they had not perceived before. Research leaders were very interested in how their faculty performed against other faculties. This is a powerful way to make the RDM agenda concrete for research managers. It led to concrete action, such as invitations to present to research groups and projects, increased requests for advice and guidance on research ethics and data discoverability, involvement with big data initiatives and the development of pilots and use cases around issues such as electronic laboratory notebooks (eLNS) and data management plans. In terms of change management and ensuring that RDM remains a strategic priority within the institution, the DAF survey results proved to be a critical resource.

One of the purposes of the paper was to explore comparing the Sheffield results to those of other institutions. Yet making more than very generalised comparisons across institutions from published data is not straight forward. There are a number of reasons for this.

Firstly, it is hard to compare the institutions concerned because of fundamental institutional differences. It's problematic to compare the results from less research intensive institutions because of the different scale of research and nature of research culture and practices. If we assume that practices across disciplines is a significant variable, the balance of subjects studied between institutions researched makes comparison difficult. Furthermore, most studies have been analysed at faculty level, not departmental level, presumably because of the amount of data available, but faculties are composed differently in different institutions.

Secondly, while most surveys report the number of responses per faculty they do not report what the total population in that group who could have answered were, so we do not know the response rate and therefore it is unclear how representative replies are. The balance of roles of participants is also a significant variable: PhD students would be expected to answer differently from professors. 46% of those completing the Exeter survey were PhD students, whereas in other surveys this is a much lower proportion, e.g. 20% at Nottingham and 7% at Leeds. Again, most studies do not compare the numbers responding by role to the total numbers in the institution/faculty by role, to calculate an actual response rate.

A third issue is that since the raw data is not available for comparison we only have figures as reported and sometimes charts presented in the published versions do not give precise results.

A fourth but very important issue is that although using the umbrella of DAF, there is significant variability in the questions asked and the wording of questions in the different surveys. Table 1 above illustrates the differences just in the numbers of questions asked, for example. Inevitably this makes direct comparison between survey results complex. Questions about data practices are more standard, but questions about data sharing, for example, are diverse. An analysis of the surveys shows that they asked between one and nine questions about data sharing (average about five). Surveys always ask about data sharing, but the questions have been quite dispersed, with perhaps eleven categories of questions:

1. Have you used data from external sources?

2. Who can typically access data you create?

3. Do you share data during or after project? With whom and how?

4. What are the reasons not to share? Or to share?

5. Who owns your data?

6. Does your data have potential for reuse for research or training?

7. Are you willing to share data? Do you plan to share data?

8. Have you deposited data before in a national subject data archive?

9. Would you be willing to use an institutional repository if it existed?

10. Are there funder requirements that you are aware of?

11. Has a publisher ever asked you to share data?

Even where the same broad question is being asked, the wording is often different or qualified in significant ways that makes direct comparison problematic. The authors of the Sheffield survey themselves inadvertently reinforced the comparability problem by introducing questions from Tenopir et al.'s (2011) study! Looking at how these questions were answered in the earlier survey, there is the further issue that this was conducted 2009/10; in that time attitudes are likely to have changed. Nevertheless, the questions asked by Tenopir et al. are important reference points for comparison, and though not published at the time of writing the study was repeated in 2013. Some comparisons can be made. Interestingly, there was lower agreement to statements about lack of access to data either holding back progress of research in general or the individuals' ability to answer their own research questions. 67% agreed or strongly agreed with the first statement in the Tenopir study; 50% to the second. This compared to 43 and 31 percent at Sheffield. The same sort of message of less commitment to data

sharing emerges from a comparison of the results presented in Table 4. All the results suggest that the type of person motivated to fill in a cross-institutional study may not represent typical views within institutions.

As further illustration of the issue around comparability. Table 5 presents an attempt to compare findings on training needs. Comparison can only be made on limited basis. The figures from the University of Nottingham are presented in graph form, hence percentages are approximate. Significantly different questions were asked. Starred questions were not asked at Sheffield. Questions were worded slightly differently, with questions asked by the University of Essex markedly different from the other two. For example, Essex's third placed item tied ethical and legal aspects together and this received one of the higher scores. Nottingham and Sheffield both had a question purely about ethics, this rated quite low. In the Sheffield survey respondents could say that they were definitely or might be interested in training; whereas in the other surveys it was a yes/no choice. This could explain why the figure of having any interest in training is consistently much higher at Sheffield. With those provisos, it does appear that data management planning and storing data are consistently rated highly across all three studies. This would support the idea that there are universal areas of interest among researchers.

Table 5. Training needs as reported in three studies (starred questions were not asked at Sheffield).

Answer rank	Sheffield	Nottingham	Essex
Top	Developing a research data management plan (74%)	Developing a research data management plan (c. 60%)	Planning for data management and sharing (50%)
Second	Storing your research data (72%)	Storing your research data (c. 46%)	Storage and back up of data and files (44%)
Third	Documenting your research (72%)	Copyright and Intellectual Property (c. 40%)	Ethical and legal aspects of data sharing and reuse (34%)
Fourth	Sharing your research data (71%)	Documenting your data (c. 38%)	Costing data management, preservation and sharing* (32%)
Fifth	Copyright and Intellectual Property (70%)	Sharing your data (c. 29%)	Freedom of information and research data* (28%)
		Data repositories and open access* (c. 29%)	Data copyright and intellectual property (28%)

In some areas more comparable questions were asked across the surveys, e.g. at the level of data practices. Looking at where researchers stored data, Sheffield researchers named on average around 4.0 places, whereas in the Nottingham survey this seems to have been more like 4.6. 39% of Sheffield researchers had less than 50GB of data; the figure was 35% at Nottingham. At Sheffield 28% backed up data daily; 17% weekly. At Nottingham the figures were 35% and 16% respectively. These types of figures suggest common patterns at this basic level.

Conclusion

The main purpose of a DAF-based survey is as a tool for institutional change and RDM user requirements gathering to inform service delivery and infrastructure. Sending the survey out in itself alerts researchers to the institutional importance of RDM. The survey results can be used in strategy building, advocacy, identifying priority areas for service delivery (technical and socio-technical support infrastructures), and for engaging other key RDM stakeholders, such as the library, research office, and IT staff. As such it is logical to adapt the approach to local requirements. Nevertheless, the value of such surveys is enhanced when comparable data is generated. It is of great value for institutions to benchmark their performance against comparable institutions. The divergence of questions and question wording in surveys is an obstacle to this, as is the failure to fully report response rates. Given the topic, in surprisingly few cases have the original data been shared, which would also increase comparability. Quite a few local studies have been done but not been published. If, as a community, we shared data more fully and were careful about the choice of question wording we could help to build a clearer picture across the sector. There would still remain barriers to comparability in terms of different institutional profiles, but we could start to identify common problem areas, such as disciplinary fields with particular problems, leading to more joint services and working.

References

Akers, K.G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation, 8*(2) 5-26. doi:10.2218/ijdc.v8i2.263

Alexogiannopoulos, E., McKenney, S., & Pickton, M. (2010). *Research data management project: A DAF investigation of research data management practices at The University of Northampton.* Retrieved from http://nectar.northampton.ac.uk/2736/1/Alexogiannopoulos20102736.pdf

Grace, S. (2013). *Results of University of East London survey on managing research data.*

HATII. (2009). *Data Asset Framework implementation guide.* Retrieved from http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf

Jones, S., Ball, A., & Ekmekcioglu, Ç. (2008). The data audit framework: A first step in the data management challenge. *International Journal of Digital Curation, 3*(2), 112-120. doi:10.2218/ijdc.v3i2.62

Jones, S., Ross, S., and Ruusalepp, R. (2009). *Data Audit Framework methodology.* Glasgow, UK: HATII. Retrieved from http://www.data-audit.eu/DAF_Methodology.pdf

Knight, G. (2013). *Research data management at London School of Hygiene and Tropical Medicine: Web survey report.* Retrieved from http://blogs.lshtm.ac.uk/rdmss/files/2013/04/LSHTM-RDM-Web-Survey-Report.pdf

Lewis, M.J. (2010). Libraries and the management of research data. In S. McKnight (Ed), *Envisioning Future Academic Library Services* (pp. 145–168). London: Facet.

Nassiri, S., & Worthington, B. (2012). *Digital Asset Framework (DAF) survey analysis.* Retrieved from http://research-data-toolkit.herts.ac.uk/document/rdtk-data-asset-survey-digest-july-2012/

Open Exeter Project Team. (2012). *Summary findings of Open Exeter DAF survey.* Retrieved from https://ore.exeter.ac.uk/repository/handle/10036/3689

Parsons, T., Grimshaw, S., & Williamson, L. (2013). *Research data management survey, the University of Nottingham.* Retrieved from http://admire.jiscinvolve.org/wp/files/2013/02/ADMIRe-Survey-Results-and-Analysis-2013.pdf

RoaDMaP. (2013). *University of Leeds research data survey 2012* [Data set]. University of Leeds. Retrieved from http://library.leeds.ac.uk/info/377/roadmap/122/roadmap_project_outputs/7

Rolando, L., Doty, C., Hagenmaier, W., Valk, A., & Parham, S.W. (2013). *Institutional readiness for data stewardship: Findings and recommendations from the Georgia Tech research data assessment.* Retrieved from http://hdl.handle.net/1853/48188

Simukovic, E., Kindling, M., & Schirmbacher, P. (2014). Unveiling research data stocks: A case of Humboldt-Universitaet zu Berlin. In *iConference 2014 Proceedings* (pp. 742–748). doi:10.9776/14351

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A,U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PloS ONE, 6*(6), e21101. doi:10.1371/journal.pone.0021101

University of Hertfordshire. (2012). *Research data assets survey.* Retrieved from http://research-data-toolkit.herts.ac.uk/document/rdtk-data-asset-survey-results-jul-2012/

University of Lincoln. (2012). *Research data management survey.* Retrieved from http://orbital.blogs.lincoln.ac.uk/files/2012/04/DAF-summary-anonymised.pdf

University of Southampton. (2011). *Data management questionnaire results: IDMB Project.* Retrieved from http://eprints.soton.ac.uk/196243/1/IDMB_Survey_Report.pdf

Van den Eynden,V., Ensom,T., & Corti, L. (2013a, March 24). *University of Essex staff research data survey*. UK Data Archive, University of Essex. Retrieved from http://www.data-archive.ac.uk/media/391114/rdessex_staffsurveyreport.pdf

Van den Eynden,V., Ensom,T., & Corti, L. (2013b, March 25). *University of Essex staff research data survey* [Data sheet]. UK Data Archive, University of Essex. Retrieved from http://www.data-archive.ac.uk/media/391120/rdessex_staffsurveyresults.pdf

Whyte, A. (2014). A pathway to sustainable research data services from scoping to sustainability. In G. Pryor, S. Jones, & A. Whyte (Eds.), *Delivering research data management services: fundamentals of good practice* (pp. 59–88). London: Facet.

Wilson, J. (2013). University of Oxford research data management survey 2012: The results [Web log post]. Retrieved from http://blogs.it.ox.ac.uk/damaro/2013/01/03/university-of-oxford-research-data-management-survey-2012-the-results/

Wilson, J., Jeffreys, P., Patrick, M., Rumsey, S., & Jefferies, N. (2012). *Results of the 2012 University of Oxford research data management survey* [Data set]. Oxford University. Retrieved from http://damaro.oucs.ox.ac.uk/docs/OxfordRDMsurvey2012_public.xlsx

Witt, M., Carlson, J., Brandt, D.S., & Cragin, M.H. (2009). Constructing data curation profiles. *International Journal of Digital Curation, 4*(3), 93-103. doi:10.2218/ijdc.v4i3.117

Appendix 1
The University of Sheffield Research Data Management Survey

About You

The purpose of this first section is to gather information on the survey respondents. If you would like to discuss your Research Data Management (RDM) and sharing needs with the RDM Support team then please answer questions one and two (the optional questions). Information entered here will not be used for any other purpose and all responses will be anonymised.

1. Please provide your name (optional)

2. Please provide your email address (optional)

3. Which of the Faculties does your research fall under?

4. Which academic department do you work for? If you are not affiliated with a particular department or faculty then please indicate this and give the academic discipline which you think best fits your area of research.

5. Which of the following best describes your job title/research role?

About Your Research and Your Research Data

In this section we would like to find out about your current research and how you create and manage your research data. Please select one or more funded or unfunded research projects that you are currently working upon.

1. Who, if anyone, funds the research project (or projects) that you are currently undertaking?

2. Do you conduct your research as part of a team or as an individual?

3. What types of research data do you create or work with as part of your research?

4. Where do you store your research data during the life of your project/research activity?

5. How many years are you required to retain your research data for after the end of a project?

6. Are there legal, regulatory, or confidentiality issues that influence how your Research Data is stored, managed and shared?

7. Please estimate the volume of research data you are expecting to create during your current research activities.

8. How frequently do you back-up your research data?

RDM Awareness and Training

This section of the survey contains questions related to RDM awareness of institutional, funder, and publisher expectations regarding research data and training requirements.

1. Before reading this questionnaire, were you aware of The University of Sheffield Research Data Management Policy? See http://www.shef.ac.uk/ris/other/gov-ethics/grippolicy/practices/all/rdmpolicy

2. Have you received any information previously from the University relating to research data management (for example via training, induction sessions, online support material, workshops)?

3. Did you develop a research data management plan (DMP) for your research?

4. Are you aware of any research data management requirements or expectations from your funder?

5. Please rate your interest in training on the following RDM topics:
   1. Funders requirements and research data management
   2. Developing a research data management plan
   3. Documenting your research data
   4. Storing your research data
   5. Creating metadata for research data
   6. Ethics and consent
   7. Sharing your research data
   8. Copyright and Intellectual Property Right (IP)

9. Citing your research data

6. Are there other areas of training related to RDM that you require?

7. Do you plan to make your research data publicly discoverable and accessible following completion of your research project/activity?

8. Have you, or your research team, ever deposited your research data in a dedicated subject/disciplinary repository?

9. Tell us how much you agree with each statement:

   1. Lack of access to data generated by other researchers or institutions is a major impediment to progress in my subject discipline.

   2. Lack of access to data generated by other researchers or institutions has restricted my ability to answer research questions.

   3. Data may be misinterpreted due to complexity of the data.

   4. Data may be misinterpreted due to poor quality of the data.

   5. Data may be used in other ways than intended.

10. Tell us how much you agree with each statement:

   1. I would use other researchers' datasets if their datasets were easily accessible.

   2. I would be willing to place at least some of my data into a central data repository with no restrictions.

   3. I would be willing to place all of my data into a central data repository with no restrictions.

   4. I would be more likely to make my data available if I could place conditions on access.

   5. I am satisfied with my ability to integrate data from disparate sources to address research questions.

   6. I would be willing to share data across a broad group of researchers.

   7. It is important that my data is cited when used by other researchers.

   8. It is appropriate to create new datasets from shared data.