

Research Data Management Practices: Synergies and Discords Between Researchers and Institutions

Sally Vanden-Hehir
The University of Edinburgh

Helena Cousijn
Elsevier

Hesham Attalla
Elsevier

Abstract

The aim of this study was to explore the synergies and discords in attitudes towards research data management (RDM) drivers and barriers for both researchers and institutions. Previous work has studied RDM from a single perspective, but not compared researchers' and institutions' perspectives. We carried out qualitative interviews with researchers as well as institutional representatives to identify drivers and barriers, and to explore synergies and discords of both towards RDM. We mapped these to a data lifecycle model and found that the contradictions occur at early stages in the lifecycle of data and the synergies occur at the later stages. This means that for future successful RDM, the points of discord at the start of the data lifecycle must be overcome. Finally, we conclude by proposing key recommendations that could help institutions when addressing both researcher and institutional RDM needs.

Received 05 July 2017 ~ Revision Received 18 April 2018 ~ Accepted 18 April 2018

Correspondence should be addressed to Hesham Attalla, Radarweg 29, 1043 NX Amsterdam, The Netherlands.
Email: h.attalla@elsevier.com

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution Licence, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introduction

Research data management and sharing is a widely discussed topic which is gaining the interest of researchers, institutions and funders alike. It is considered to be an important aspect of academic research, and central to research reproducibility and scientific integrity. Many funding agencies have adopted policies requiring compliance with data management and sharing plans. The National Institute of Health states that ‘data should be made as widely and freely available as possible’ (NIH, 2003), and the Research Councils UK stipulate that publicly funded research should be open to the public (RCUK, 2011). The National Science Foundation’s policy requires a data management plan to be submitted with grant proposals (National Science Foundation, 2014), and the EU funded Horizon 2020 project instructs that data should be ‘as open as possible, as closed as necessary’ (European Commission, 2016). Increasingly, these policies draw on the FAIR Data Principles, which state that research data should be findable, accessible, interoperable and reusable (Wilkinson, 2016).

Researcher incentives for RDM have been previously documented, with one study concluding that sharing a dataset with an article leads to an 85% increase in citation rate over an article with no supporting data (Piwowar, Day and Fridsma, 2007). This was validated with a follow up study which gave a less pronounced effect but used a larger sample size (Piwowar and Vision, 2013). Data sharing can also create opportunities for collaboration and the faster advancement of science (Goodman et al., 2014; Sommer, 2010; Whyte and Pryor, 2011). As scientific research becomes increasingly collaborative and interdisciplinary, a study correlated the increase in number of authors on a paper with increasing impact of this paper, giving a driver for collaboration (Wuchty, Jones and Uzzi, 2007).

The main researcher barriers towards RDM previously reported were the extra time taken to prepare datasets for archiving and sharing, and the fear of data misuse. Several studies found that the time and effort needed to prepare data was discouraging researchers from sharing their data (Fecher, Friesike and Hebing, 2015; Kim and Stanton, 2016). One survey reported that for 80% of respondents, the leading barrier to RDM was the fear of not establishing priority for their work if other researchers used their data to publish before them (Fecher, Friesike, Hebing, Linek and Sauermann, 2015).

However, attitudes to data sharing also vary with levels of researcher seniority, and between subject disciplines. Younger scientists are less willing to share their data in a public repository than scientists aged over 50 (Tenopir et al., 2011). A study that grouped participants into the disciplines of arts and humanities, social sciences, medical sciences and basic sciences (biology, chemistry and physics) reported that the basic scientists are most likely to share their data (Akers and Doty, 2013). The most common reason to withhold data for medical scientists was sensitive data, while for basic scientists it was fear of not being properly recognised. However, due to the limitations of qualitative research, it is difficult to include enough information about participants to provide context for interpretation, whilst still protecting participants’ anonymity (Kaiser, 2009).

Data management has been well established for some time in certain fields, such as genomics, astronomy and physics, which routinely produce large amounts of data and rely on collaboration. The current challenge is to convince researchers who do not

produce large amounts of data, described by Borgman as ‘the long tail’ of research, to adhere to sound RDM practices (Borgman et al., 2016). Small datasets make up the majority of academic research, and are typically stored and maintained privately. These are the type of datasets that academic librarians will most likely be required to help manage (Akers, 2013).

There are limited previous studies which consider RDM from an institutional perspective, but there have been some studies investigating library involvement and perceptions. A survey of librarians in the UK reported that 31% of institutions had a RDM policy in place, and that in most cases the library had been involved in setting it up (Cox and Pinfield, 2013). Moreover, 70% of the respondents also reported that they felt a cultural change towards more stringent RDM in their institutions. The same authors carried out semi-structured interviews to create a model of institutional RDM encompassing drivers, programme components, influencing factors and stakeholders (Pinfield, Cox and Smith, 2014). They indicated that the main institutional drivers for RDM are storage, security, preservation, compliance, quality, sharing and jurisdiction.

Although perceptions, drivers and barriers for RDM have been reported for both researchers and librarians, there have been no previous studies, to the best of our knowledge, which combine and compare drivers and barriers both for researchers and institutions at the same time. We believe that this study uncovers important synergies and discords between researchers and institutions when it comes to RDM practices. Understanding the dynamics of interactions between these factors can help institutions when designing the infrastructure aimed at supporting their researchers’ workflow.

Study Aims

Through two rounds of qualitative interviews with 1) researchers and 2) institutional representatives, we aimed to answer the following questions:

- What are the RDM drivers and barriers for researchers and institutions?
- Are there any synergies or discords of RDM drivers and barriers between these two groups?
- At which points in the research data lifecycle do these synergies and discords arise?

In the results section we report the drivers and barriers found from our analysis of the interview transcripts. In the discussion we show where there are synergies and discords between the researchers and institutions, and where they arise on a data lifecycle model.

Methods

We carried out two rounds of semi-structured interviews between July and December 2016, firstly with academic researchers (n=14) of varying levels of seniority and working at different universities, (see Appendix), and secondly with institutional representatives (n=12), who are responsible for RDM at their respective universities, (see Appendix). The institutions were chosen mainly from Europe and North America as this is where RDM seems to be most well established. The interviews were mainly

carried out with RDM librarians; we recognise that this could be a limitation of the study as the views expressed may reflect those who implement RDM policy but not those who design it, which in many cases is driven by the office of research or equivalent. We chose to carry out semi-structured interviews as they allow the same ideas to be discussed with all participants, whilst giving the participant and interviewer the freedom to achieve more depth of personal experiences (Gill, Stewart, Treasure and Chadwick, 2008).

The participants for the researcher interviews were all recruited from a pool of researchers representing different demographics, disciplines and seniority levels, and the institutional representatives were recruited by sending personal invitations to individuals known to be working with RDM in their institutions. This sample was chosen to cover a range of researchers and research institutions who are concerned with RDM; we stopped recruiting further participants when we felt that saturation of the responses was achieved. The interview script was designed to explore what motivates both researchers and institutional representatives to properly manage their research data, what prevents them from proper management, the risks associated with that, and the resources available to them. All interviews were carried out remotely using the videoconference tool WebEx, were recorded after taking participants' permission and later transcribed verbatim; each interview lasted approximately 60 minutes. Transcriptions were coded and analysed using QDA Miner Lite software. Quotes within the text are quoted verbatim unless it could reveal the identity of the respondent, in which case square brackets are used to replace the identity.

To ensure the anonymity of our participants, each has been assigned a code (see Appendix). These codes will be used throughout the text to identify quotes.

Results

The results are split into the two rounds of interviews; the first round giving researcher perspectives and the second round giving institutional perspectives. Data analysis revealed drivers and barriers for three distinctive sub categories of RDM; data storage, data sharing and data reuse.

Interview Round One: Researcher Perspectives

Drivers for data storage

The most frequently mentioned driver by researchers for data storage was having the peace of mind that their data was preserved and protected. Effective storage and backups are essential to prevent data loss, which some respondents had previously experienced causing them to maintain a habit of regular backups.

'I have three copies; computer, external hard drive and in [institutional] drive. I happen to lose the data once so now I make three copies' (R-4).

Researchers store data on the institutional drive for its automatic daily backup feature (R-3), and for security reasons (R-12). It is now common that researchers collect datasets which are too large to be stored on their personal computers. One participant, (R-4), commented that cloud storage was very useful because it can accommodate large

amounts of data, is easy to access and eliminates the need for portable external hard drives.

Strict rules from the funding agency to archive all data twice a year was the driver for another participant (R-2). Although most researchers did not see funder compliance as a strong driver for data storage currently, some recognised that this might change soon.

Barriers for data storage

A barrier reported by R-12 was that it can be difficult and time-consuming to organise and consolidate different file versions stored in different locations.

‘I always have four or five copies of the data. Usually [I] have a raw copy on the server at work, then on my laptop and one on my external hard drive. Occasionally [I] get around to consolidating them’ (R-12).

There was also a feeling of frustration that time had to be spent organising and archiving data instead of carrying out research (R-5), and another participant, (R-7), complained of a lack of storage space on their computer.

Drivers for data sharing

Researchers recognised that data sharing could increase their research impact and improve their career prospects. However, none of the participants had actually shared data, so these quotes are merely speculative. Four participants, (R-3, R-9, R-10 and R-12), reported that it would be useful for others in their field to be able to scrutinise their data, and potentially extract additional information out of it, which could lead to an increase in collaboration opportunities.

‘Opening the data would open up collaborations in the future’ (R-10).

There was a positive feeling that sharing data would move science faster by helping peers avoid repeating the same mistakes by opening up data.

‘We produce a lot of data, we only publish a little bit... Negative results are good to let people know what not to do’ (R-11).

Barriers for data sharing

The main barrier to data sharing, with over half of the participants mentioning it, was that preparing and structuring data into a format that can be easily disseminated takes too much time and effort. Participants indicated that while understanding their own data is easy, there is a need to provide additional metadata and explanations before expecting someone else to fully understand it.

‘I would be concerned... that there is clear documentation with it’ (R-12).

Another major barrier, voiced by six participants, was the fear that their data may be used without proper credit being given, especially unpublished work.

‘There is always the fear of someone using your data’ (R-13).

Others were reluctant to share as they wanted to make further analyses on the same dataset, and five researchers worked with sensitive data, and were unsure what kind of anonymization was needed before sharing.

Drivers for data reuse

The main driver for data reuse is the potential to save time by preventing the researcher from wasting time repeating unsuccessful experiments. Validation of preliminary results to those performed by more established researchers was also an important driver.

‘When you compare results in a similar area – there is a protocol – how does my control data compare to their control data?’ (R-3).

In total, five participants reported saving time by not repeating failed experiments, and comparing results with others as drivers for data reuse.

Barriers for data reuse

The lack of clear and structured metadata, and the time taken to understand another’s data comprised almost all of the barriers reported for data reuse. Also, data were often not presented in a reusable form, especially if it had been collected using a specialised instrument or proprietary software incompatible with others. Understanding the data sometimes required input from the data originator, which is also time-consuming.

‘I sometimes have questions for the data collector and it can take one-two weeks for an answer’ (R-9).

Table 1 gives a summary of all the researcher drivers and barriers to RDM.

Table 1. Summary of researcher drivers and barriers to RDM.

Drivers	Barriers
Preservation of data	Time to consolidate data
Storage of large datasets	Lack of storage space
Funder compliance	Fear of data misuse
Increase research impact	Data needed for further analyses
Move science faster	Sensitive data
Validation of results	Lack of metadata

Interview Round Two: Institutional Perspectives

Drivers for data storage

For institutions, compliance with funders’ requirements was the most frequently cited driver for data storage.

‘At an institutional level, the policy itself really says that researchers should be complying with requirements set down by the funders’ (I-1).

The archiving of completed data was also considered a driver, so that data would not be lost once a researcher leaves the institution. Institutions are eager to provide proper data management training and assistance at an early stage in research projects so that the data produced becomes an asset to the institution.

‘[The institution is] producing all this data and it seems imperative on us to be able to assist researchers and to be able to store that in a way that makes sense’ (I-2).

Many of the institutions had internal repositories to provide an archive for research data that does not fit anywhere else, as for the majority of researchers there is not an established repository for their field.

Barriers for data storage

A barrier for almost half of the institutions was that they lacked the storage space, and that it would not be feasible for them to archive all of the institution’s research data with their current funds.

‘Researchers can now gather a lot more information in a shorter time, but then all of a sudden storage becomes a problem’ (I-9).

Storing data with adequate security was another barrier, especially in the case of third party cloud storage and sensitive data.

A further barrier reported by half of the institutions was that their current RDM systems were disjointed. Individual faculties or research groups were often responsible for their own data management, making it hard for the library to gain a complete overview.

A lack of enforcement of policies by the funder was a barrier to be overcome by the institutions. I-3 reported that ‘it doesn’t seem like funders have decided what their compliance requirements are going to be’, and I-11 reported ‘There’s no way we can enforce this.’

Drivers for data sharing

In addition to data storage, institutions also want to make it easy for their researchers to make their data discoverable, so as to increase the research impact of the institution.

There was recognition that data sharing would increase the reputation of the institution, but that researchers had to be encouraged to share their data. Examples of this encouragement were to provide download statistics for datasets, (I-8 and I-11), and to link shared datasets to research articles, (I-3 and I-8). Compliance to ‘meet open data mandates’ (I-2) was also mentioned in terms of data sharing, although not as frequently as data storage.

Barriers for data sharing

Confidentiality and privacy issues were the major barrier to sharing, with institutions acknowledging that not all data can be shared.

Institution 9 believed that a cultural change was needed amongst its researchers, especially in older researchers who were not used to digital methods.

‘It is the older researchers who struggle. They have never been into digital data and archiving’ (I-9).

Drivers for data reuse

Data reuse and data citations will again increase the impact of the institution, but only if the data is in a reusable form. Reproducible research is also a highly important driver for the reputation of the institution.

‘...I think it’s probably making sure your data... your research can be replicated; can be verified’ (I-11).

I-1 also recognised the economic benefits of data reuse and commented that they would like their researchers to retain data that could be of use to others in the future.

Barriers for data reuse

The lack of metadata on datasets was a barrier to reuse. A lack of metadata standards between institutions and countries means that sometimes data can’t be reused even if it has been properly described.

A further barrier to reuse is that data in institutional repositories are not as discoverable as data in well-known subject based repositories.

‘...they prefer the subject based repositories simply because that’s where people go to look for data at this point. They generally don’t go to an institutional repository to find data...’ (I-8)

The institutional drivers and barriers towards RDM are summarised in Table 2 below.

Table 2. Summary of institutional drivers and barriers to RDM.

Drivers	Barriers
Funder compliance	Lack of storage space
Data as an asset	Inadequate security
Preservation of data	Disjointed RDM systems
Increase research impact by sharing	Lack of enforcement
Host small data	Confidentiality issues
Economic benefits of reuse	Lack of metadata

Discussion

Synergies and Discords

Our results identified drivers and barriers for both researchers and institutions in the sub categories of data storage, data sharing and data reuse, where some synergies and discords between the researchers and institutions are apparent.

Synergies

Comparing researchers' drivers to those of the institutions, we recognised that some of them aligned well, creating synergy between the two groups (see Table 3). Both are in agreement with regard to the importance and necessity of proper data preservation. Researchers store and backup their data to have peace of mind and avoid data loss, while institutions increasingly see data as another digital asset that needs curation. Moreover, both researchers and institutions are interested in increasing the impact of their research by sharing data. Taken into consideration that there is a correlation between article age and the availability of its supporting data, with the latter falling by 17% per year after publication (Vines et al., 2014), and that papers linked to data receive on average 50% more citations, (Dorch, 2012), it becomes apparent why both groups have an invested interest in proper data storage and sharing.

Table 3. Areas where the researchers and institutes are in synergy with drivers.

Synergy Area (Drivers)	Researcher	Institution
Preservation	Safely store and preserve data	Archive data in the institution
Impact	Data sharing increases individual researcher impact	Data sharing increases institutional impact
Savings	Save time by reusing data	Save money by reusing data

Researchers consider the potential increase in collaborations as a driver for data sharing, which is encouraged by the institutions. The benefits of data reuse are also shared, with researchers seeing the value in terms of saving time, while institutions see the value of saving money. This economic benefit was discussed previously by Whyte and Pryor (2011).

Researchers and institutions do not only share the same RDM drivers, but also some of the barriers for RDM are common between the two groups (see Table 4). Current RDM systems are perceived as disjointed by both groups, who agree that a centralised system would be ideal. Researchers expressed frustration over the extra time spent on RDM, preventing them from spending time on their research. This concern was shared by institutions, which are, in most cases, investing in user friendly solutions to support their researchers with RDM.

Table 4. Areas where the researchers and institutes are in synergy with barriers.

Synergy Area (Barriers)	Researcher	Institution
Fragmented institutional RDM	Confusion over multiple storage locations	RDM systems are disjointed
Lack of RDM education	Researchers are reluctant to share data	Cultural change needed within researchers
Sensitive data	Ethical issues	Legal issues
Metadata standards	Lack of standards between laboratories	Lack of standards between institutes/countries

However, not all researchers are open to data sharing; some reported that they are reluctant to share data if it has potential in another study. This is supported by a previous study where authors in PLoS journals were contacted and requested to share their data, which is a requirement of the journal's data policy (Savage and Vickers, 2009). They study found that only one out of ten researchers shared their data, with one reason for withholding being the potential for future analyses on the data. Institutions recognised this limitation and commented that a cultural change was needed among their researchers.

Dealing with sensitive data is also a concern for both groups. Researchers are concerned over the ethics of sharing data, and their inability to recruit participants if they have to openly share data. Similarly, institutions are concerned about the legal issues with sharing sensitive data. However, only a handful of the institutions we interviewed offer support to their researchers in dealing with sensitive data.

Both also agree that a lack of metadata standards is a barrier to data reuse. A lack of documentation and metadata was discussed in a previous study as a barrier for meta-analyses (Howe et al., 2013).

Discords

Researchers and institutions are not always in agreement with regard to their RDM drivers and barriers, as we illustrate in Table 5. One area of disagreement is the driver of funder compliance. Whilst some researchers are aware that funder compliance will be important in the future, it is not currently considered a strong driver for data storage. This is supported by previous studies which found that mandates and pressure from funders to share data did not always encourage researchers to do so (Kim and Stanton, 2016; Piwovar, 2011). In contrast, funder compliance was the most frequently mentioned driver from the institutions.

Table 5. Areas where the researchers and institutes are in discord.

Discord Area	Researcher	Institution
Funder compliance	Compliance low on priority list	Compliance is main driver
Cloud storage	Cloud storage is useful	Cloud storage is insecure
Metadata	Time and effort to add metadata	Encourage addition of metadata

Using third party cloud storage solutions to store research data is another point of discord. Researchers were generally positive about using cloud storage, citing the ability to store and share large datasets, and the convenience of accessibility as the main reasons to use it. In contrast, institutions discourage the use of cloud storage due to concerns over data security.

Another seeming discord is related to adding metadata to datasets. Institutions expressed a need to encourage researchers to add metadata to ensure data is easily understandable and reusable. However, the main barrier to data sharing for researchers was the time and effort taken to prepare the metadata. Some researchers see the addition of metadata as a waste of time, especially when there is no enforcement or guaranteed rewards. A previous study described academia as a ‘reputation economy’ where researchers are unlikely to adopt RDM practises unless they see reward for themselves (Fecher, Friesike, Hebing, Linek and Sauermann, 2015; Friesike, Fecher, Hebing and Linek, 2015).

Synergies and Contradictions Mapped to a Data Lifecycle Model

RDM concerns the management of data from the conceptualisation of a project, through the collection and analysis of data, preparing data for publication, and to archiving and potential sharing for future reuse. Data lifecycle models are often used to convey the different management needs at different points in the data journey (Higgins, 2008; Surkis and Read, 2015).

In Figure 1 below, we have mapped the synergies and discords identified above onto a data lifecycle model we developed. This mapping serves the purpose of explicitly identifying where researchers and institutions are currently working in synergy or in discord. The model we have created serves as a generic data lifecycle model, we recognise that all the steps may not apply to every subject discipline or individual researcher.

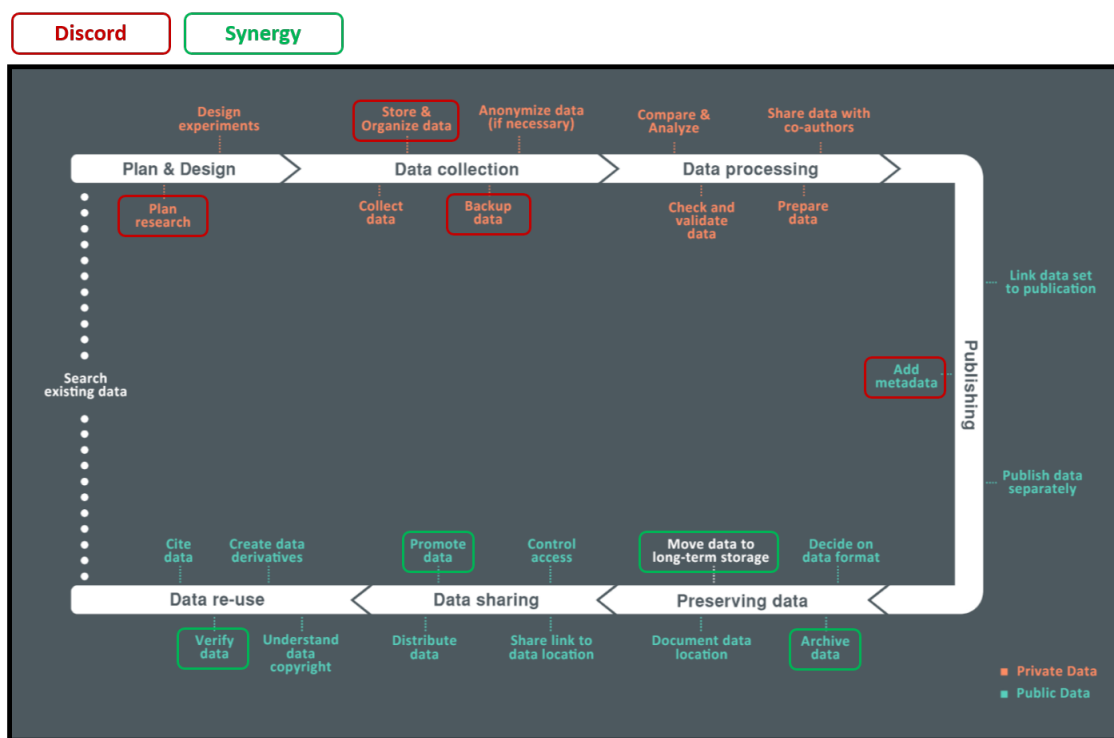


Figure 1. Synergies (green) and discords (red) mapped to a data lifecycle model.

We observe that researchers and institutions do not have the same drivers in the early phases of the data lifecycle, but are working towards the same RDM goals in the later phases. In the planning and design phase, funder compliance is an important driver for institutions, but is not effectively communicated to the researchers leading to varying levels of compliance. During the data collection phase there are conflicts over cloud storage, and whether the institution has sufficient room for backups of large datasets. The addition of comprehensive metadata is not a priority for researchers which complicates the reusability and discoverability of data later on.

In the later stages of the model, institutions and researchers are both working towards preservation and data archiving and agree on the advantages of data sharing. For the researcher, the motivation is the increased exposure to collaborations and increased citations, and for the institution sharing data could increase their impact, for example in the media. There is also agreement on the benefits of data reuse to validate results and save time and money.

Recommendations for Future Research Data Management

Our analysis shows evidence that researchers and institutions are focused on RDM at different points within the data lifecycle. Although our results were based on a limited sample size, we feel that we have identified areas of synergy and discord between researchers and institutions, and that these points of discord must be overcome in order for RDM to progress.

1. **Common drivers:** We identified three areas where researchers and institutions work in synergy towards RDM in the later stages of the lifecycle model; both value the proper preservation of data, see the benefits of increasing impact by sharing and realise the benefit of reusing data to save time and money. As there

is already synergy in these areas we believe that it is easier to align both groups by focusing on these common benefits.

2. **Common barriers:** Fragmented institutional RDM systems and a lack of researcher awareness provide common barriers to RDM. Other common barriers were the absence of metadata standards, and the issues with sharing sensitive data. These common barriers are areas that are well known in the RDM community, but our findings suggest they are still prevalent in our interviewees' minds.
3. **Points of discord:** Our results also identified points of contradiction between researchers and institutions. There is a disagreement on the use of cloud storage, the storage of large datasets, the importance of compliance with funders and others' requirements, and also on the early and comprehensive use of metadata. Our findings suggest a need to pursue institutional RDM strategies, such as educating researchers on security and the benefits of data sharing and reuse, and also providing training on the proper use and capture of metadata.

Conclusions

Through qualitative interviews we have explored the drivers and barriers for RDM from both researcher and institutional perspectives, and identified where there were points of synergy and discord between the two parties. When mapped to a data lifecycle model, we found that the discords appear in the early stages of the lifecycle and the synergies appear in later stages. We concluded by discussing what can be done to overcome the common barriers and the discords between researchers and institutions. Although we acknowledged that institutions are cognizant of the challenges facing proper RDM, we believe that this study confirms that previously reported RDM drivers and barriers are still prevalent. Our model provides a basic scheme of areas that need to be addressed when designing RDM systems for institutions and researchers alike.

References

- Akers, K.G. (2013). Looking out for the little guy: Small data curation. *Bulletin of the American Society for Information Science and Technology*, 58–59. Retrieved from http://www.asis.org/Bulletin/Feb-13/FebMar13_RDAP_Akers.pdf
- Akers, K.G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *The International Journal of Digital Curation*, 8(82), 5–26. doi:10.2218/ijdc.v8i2.263
- Borgman, C.L., Golshan, M.S., Sands, A.E., Wallis, J.C., Cummings, R.L., Darch, P., & Randies, B.M. (2016). Data management in the long tail: Science, software, and service. *International Journal of Digital Curation*, 11(1), 128–149. doi:10.2218/ijdc.v11i1.428

- Cox, A.M., & Pinfield, S. (2013). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 1–18. doi:10.1177/0961000613492542
- Dorch, B. (2012). On the citation advantage of linking to data: Astrophysics. *HAL Archives*. Retrieved from <https://hal-hprints.archives-ouvertes.fr/hprints-00714715v2/document>
- Faniel, I.M., & Silipigni Connaway, L. (2017). Librarians' perspectives on the factors influencing research data management programs. *College & Research Libraries Pre-Print*. doi:10.5860/crl.79.1.100
- Fecher, B., Friesike, S., Hebing, M., Linek, S., & Sauermann, A. (2015). A reputation economy: Results from an empirical study on academic data sharing. *DIW Berlin, February*, 1–26. doi:10.2139/ssrn.2568693
- Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PLoS ONE*, 10(2).
- Friesike, S., Fecher, B., Hebing, M., & Linek, S. (2015). Reputation instead of obligation : Why we need to forge new policies to motivate academic data sharing. Blog Post. Alexander von Humboldt Institute for Internet and Society. Retrieved from <http://www.hiig.de/en/23202/>
- Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal*, 204(6), 291–295. doi:10.1038/bdj.2008.192
- Goodman, A., Pepe, A., Blocker, A.W., Borgman, C.L., Cranmer, K., Crosas, M., ... Slavkovic, A. (2014). Ten simple rules for the care and feeding of scientific data. *PLoS Computational Biology*, 10(4), e1003542. doi:10.1371/journal.pcbi.1003542
- European Commission. (2016). Guidelines on FAIR Data Management in Horizon 2020. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134–140. doi:10.1145/1378889.1378998
- Howe, G., Sperling, A., Beardslee, W., Sandler, I., Shern, D., Pantin, H., ... Brown, C. H. (2013). Advancing science through collaborative data sharing and synthesis. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, 8(4), 433–44. doi:10.1177/1745691613491579
- Kaiser, K. (2016). Protecting respondent confidentiality in qualitative research. *Qualitative Health Research*, 19(11), 1632-1641. doi:10.1177/1049732309350879

- Kim, Y., & Stanton, J.M. (2016). Institutional and individual influences on scientists' data sharing behaviours: A multilevel analysis. *Journal of the Association for Information and Science Technology*, 67(4), 776–799. doi:10.1002/asi.23424
- National Science Foundation. (2014). Chapter II - Proposal preparation instructions. Retrieved from https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#dmp
- NIH. (2003). Data sharing policy and implementation guidance. Retrieved from https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
- Pinfield, S., Cox, A.M., & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers and influences. *PLoS ONE*, 9(12), 1–28. doi:10.1371/journal.pone.0114734
- Piwovar, H.A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE*, 6(7), e18657. doi:10.1371/journal.pone.0018657
- Piwovar, H.A., Day, R.S., & Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3).
- Piwovar, H.A., & Vision, T.J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. doi:10.7717/peerj.175
- RCUK. (2011). Common principles on data policy. Retrieved from <http://www.rcuk.ac.uk/research/datapolicy/>
- Savage, C.J., & Vickers, A.J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE*, 4(9), e7078. doi:10.1371/journal.pone.0007078
- Sommer, J. (2010). The delay in sharing research data is costing lives. *Nature Medicine*, 16(7), 744–744. doi:10.1038/nm0710-744
- Surkis, A., & Read, K. (2015). Research data management. *Journal of Medical Library Association*, 103(3), 154–156. doi:10.3163/1536-5050.103.3.011
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6), e21101. doi:10.1371/journal.pone.0021101
- Vines, T.H., Albert, A.Y.K., Andrew, R.L., Débarre, F., Bock, D.G., Franklin, M.T., ... Rennison, D.J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, 24(1), 94–97. doi:10.1016/j.cub.2013.11.014
- Whyte, A., & Pryor, G. (2011). Open science in practice: Researcher perspectives and participation. *International Journal of Digital Curation*, 6(1), 199–213. doi:10.2218/ijdc.v6i1.182

Wilkinson, M.D., et al. (2016). Comment: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018, 1–9.
[doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Wuchty, S., Jones, B.F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–9.
[doi:10.1126/science.1136099](https://doi.org/10.1126/science.1136099)

Appendix

Table 6. Demographics of the researchers.

Researcher Code	Position	Location	Discipline	Research Topic
R-1	Lecturer (mid-career)	Australia	Medical & health sciences	Clinical trial analysis
R-2	Post doc (mid-career)	India	Life sciences	Nanotechnology / Food industry
R-3	Post doc (mid-career)	Mexico	Life sciences	Neuroscience
R-4	PhD student (early career)	Netherlands	Life sciences	Protein quantification
R-5	Assistant professor (mid-career)	UK	Physics	Medical physics
R-6	Assistant professor (mid-career)	Greece	Physics	Biomedical spectroscopy
R-7	Principal investigator (senior)	USA	Life sciences	Infectious diseases
R-8	Principal investigator (senior)	Germany	Materials science	Steel microscopy
R-9	PhD student (early career)	Brazil	Social sciences	Psychology
R-10	Lecturer (mid-career)	Australia	Engineering & technology	Energy consumption
R-11	Principal investigator (senior)	Spain	Physics	Magnetic nanoparticles
R-12	Research associate (mid-career)	UK	Social sciences	Heritage sector
R-13	PhD student (early career)	Taiwan	Chemistry	Nanomaterials in energy fields
R-14	Post doc (mid-career)	USA	Engineering & technology	Nanotechnology

Table 7. Demographics of the institutional representatives.

Institution Code	Position	Location
I-1	Research management	UK
I-2	RDM librarian	US
I-3	RDM librarian	Canada
I-4	IT	Germany
I-5	Research management	Italy
I-6	Research management	Italy
I-7	RDM librarian	USA
I-8	RDM librarian	USA
I-9	RDM librarian	Netherlands
I-10	RDM librarian	Austria
I-11	RDM librarian	Singapore
I-12	RDM librarian	USA