

Current Practices in Research Data Management and Sharing

February 2020

A. Abstract

1. Research Data Management (RDM) covers a wide spectrum of activities and requirements. Spurred by some funder policy requirements, institutions have begun to address the policies, processes, systems, services, behaviours and cultures that are necessary across the lifecycle of research data. Drawing on a series of informal, semi-structured interviews with a wide range of institutions undertaken on behalf of Jisc, this paper summarises the state of play of RDM, highlights the common challenges, and identifies the substantive areas for development.
2. The key observations are that:
 - The single largest driver for institutional engagement in RDM has been the Engineering and Physical Sciences Research Council's policy mandate.
 - Institutional approaches to RDM have tended to be focused on compliance, rather than on using RDM to enable research and its outcomes.
 - The nature, level and placement of institutional leadership engagement varies significantly, and is quite susceptible to changes in personnel. The topic also seems to have a relatively low institutional priority.
 - There is a range of academic engagement, both with the principles and with the practices. Some of the most innovative practices were being demonstrated by institutions with a significant creative practice presence.
 - The motivations for and benefits of RDM are not adequately articulated and demonstrated to persuade academic researchers to change their behaviours. This is even more the case with data sharing.
 - Greater efforts, in policy, incentives, mechanisms and support should be applied at the earlier stages of data capture, management and analysis, in order to address academic interests, thus gaining engagement and improving quality of data for later sharing.
 - Supporting open access to publications in order to meet the requirements of the Research Excellence Framework tends to be dominant, consuming attention and resource.
 - Few institutions reported making active use of data management plans to help them manage and maximise the use and hence value of their data sets.
 - All institutions accepted deposit in an appropriate subject repository or equivalent, but not all collected all of the metadata themselves.
 - Digital preservation is relatively nascent, typically being driven by an active special collections or archive function.
 - Services to support RDM are constrained and in some cases are fragmented. Greater attention needs to be paid to integrating the expertise required across functional and structural boundaries.

B. Introduction

3. Research Data Management (RDM) covers a wide spectrum of activities. Figure 1 illustrates the research data lifecycle, from creation and management to use. This lifecycle is decadal, and potentially multi-decadal, which creates a number of challenges. In order to make it worth storing data sets for such lengths of time requires thoughtful design, capture, analysis and curation, along with transparency to allow discoverability and usability. Looking after data is clearly not a simple task.

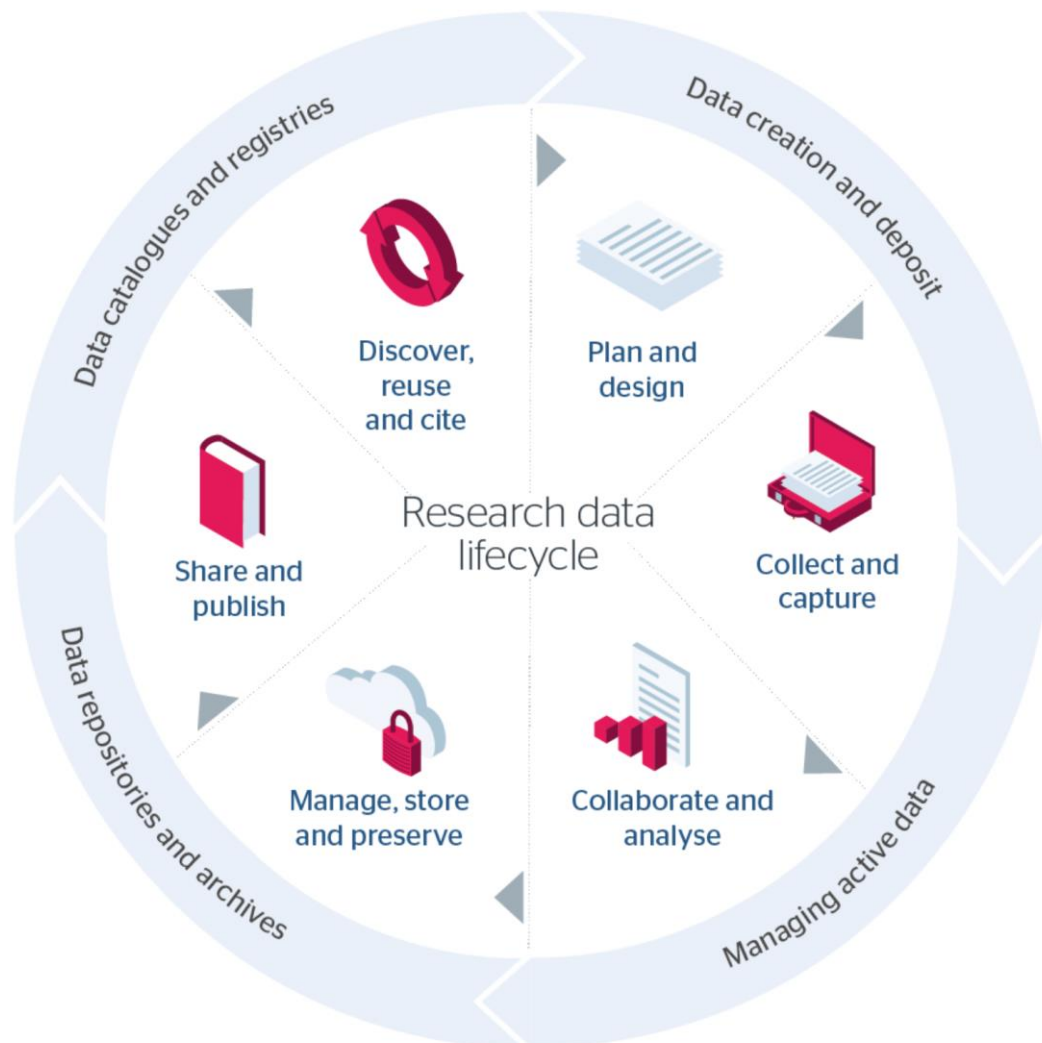


Figure 1: The Research Data Lifecycle (<https://rdmtoolkit.jisc.ac.uk/research-data-lifecycle/>)

4. In an exercise commissioned by Jisc, a range of UK institutions were selected by Jisc to be interviewed to explore their motivations, state of play, and challenges in research data management. The conversations followed a semi-structured interview format, allowing for significant variation for each institution, to reflect the topics of most relevance or interest to them. Three broad headings were used: appetite and status; practical issues; and financial and commercial considerations; with the first taking the majority of the time in the conversation. In total, 47 institutions were interviewed from late 2018 to the middle of 2019; see the Appendix for the full list. The roles of the individuals interviewed included staff from the Library, Research Services, Research Governance, IT Services, and Pro Vice-Chancellors.

5. I am grateful for the openness with which individuals discussed the opportunities and challenges that they face. Any errors in or omissions from the report are mine.

C. Observations

6. The following observations are drawn from across the interviewees. Whilst there was a significant amount of commonality, there was also an interesting amount of variation.
7. Whilst some institutions suggested that they were small or specialist or less research intense and hence would have less to say or to contribute, that was typically not the case. Indeed, some of the 'smaller' institutions demonstrated some aspects of good or innovative practice, whilst some of the 'larger' institutions were a little restricted in their thought or practice.
8. What was generally common was that those tasked with supporting RDM were doing their best in constrained circumstances. They were often isolated in being the only one or ones who understood the full import of research data and what it requires, or what is required.
9. The observations presented here are necessarily based on the people interviewed and their experiences. This is not a rigorous academic survey or study, and hence does not purport to be a wholly representative set of findings. A number of institutions who were not selected to be interviewed are known to be leaders in the field.
10. The observations have been grouped under eight headings, being common and inter-related topics: drivers; leadership; academic engagement; evidencing benefits; repositories; RDM services; digital preservation; and functionality required.

C.1 Drivers of Approaches to RDM

11. In almost all cases, the Engineering and Physical Science Research Council (EPSRC) research data policy¹ was the prime mover for an institution to have developed their capacity to support RDM. Having a reasonable period in which to respond (four years) and a defined target date (May 2015) was helpful. The effect was to spur institutions to have a suitable policy and to have a repository for data. However, in many cases, the institutional response did not go further than this. Indeed, with the EPSRC being perceived to have stepped back from advocating let alone enforcing its policy, institutions have become less motivated to have RDM high up their list of institutional or research priorities. This has left those with operational responsibility for supporting RDM somewhat isolated in many institutions.
12. The REF has had a marginal effect, as there is no current strong mandate. Some interviewees felt that having a REF mandate would strengthen their hands, in persuading both their institution to address the topic and their researchers to deposit data. However, many expressed a concern that a mandate-based approach would lead to a compliance-based response. Many also observed that RDM had not developed sufficiently (technically and culturally) to be able to support a strong mandate. Their view was that such a compliance-based response may not produce all of the desired benefits and may create tensions between the research community and those seeking to support data management. Only a small number of institutions reported making

¹ <https://epsrc.ukri.org/about/standards/researchdata/> Published in 2011, with deadlines in 2012 and 2015.

specific efforts to be able to present a developed position on open research data in the REF.

13. The drivers on researchers (as reported by the interviewees) tended to be related to subject practices, to policies applying to individual projects, and to journal requirements. Of these three, it seems, anecdotally, that funder mandates are the least likely to create positive researcher behaviours. Subject practices are about getting research done and journal requirements are about getting research published, both of which resonate strongly with researchers. Several interviewees commented that it has been their EPSRC-funded researchers who are more resistant to sharing their data. In one case it was reported that a researcher had said that they were only making their data available because it was a funder requirement to do so, but they did not intend to make their data that findable or that usable. Whilst journal requirements might drive behaviours, some journal policies were observed to be crude rather than nuanced.
14. As indicated, the general institutional response has been one of seeking to achieve compliance. There seems to be less of a driver to undertake RDM as good practice in research. In addition, there seems to be a greater focus on open data compliance than on good data management itself, when one needs decent, well-managed data to make it worth sharing. One interviewee explicitly noted that their approach was only to deposit data that was appropriately structured and documented, in comparison to another (named) institution that accepted any data sets.
15. Whilst observing a greater emphasis on data sharing, only a small number of institutions seem to have a more comprehensive ambition relating to open research / open science / open scholarship of which data sharing and open access (OA) to results are parts. It is also interesting that the Concordat on Open Research Data was only mentioned in the interviews a very small number of times, although that does not mean that institutions are not considering its principles.

C.2 Leadership of RDM

16. Discussion of the leadership of RDM within each institution highlighted significant weaknesses and fragilities. In most cases, senior institutional leadership had been involved at the outset, in response to the EPSRC policy. However, this does not seem to have been consistent or continued in many cases, for which there may be three causes. First is the belief that having a policy and a repository is sufficient; i.e. it is a technical problem that has been solved. Second is the perception that the topic has a reduced priority to some extent, given EPSRC's lack of follow-through and that RDM is not mandated in REF2021; i.e. in today's fire-fighting world, only the highest priority or most immediate topics get attention. Third is that the leadership engagement was highly dependent on the individual, and that there was a lot of change in personnel over the period being discussed; i.e. there was a lack of a long-term, consistent, senior level, institutional approach.
17. Each of these causes is interesting, and illustrates challenges and gaps. Whilst there is a lot of technical content to RDM, the potentially largest hurdle is that of research cultures and behaviours, as has been seen with Open Access to results. Pursuing a technical approach can lead to, well, a technical solution, but one that does not necessarily respond to researchers' needs or interests: researchers may be more interested in tools to help them capture, manage and analyse data, both active data and stored data. Even if a technical solution is appropriate, it will often need advocacy and support, both to introduce and over the long term, which may not get factored into the operational plans of all relevant parts of the institution.

18. The relative priority of any particular topic or policy is important to recognise. In of itself, a topic may be important and an intervention worthy. However, within an institutional context, there is a natural competition for attention, resource and time, both amongst research-related topics and in the balance between research, teaching, innovation and any other activities. Several institutions indicated that RDM did not have a high enough current priority to gain active investment, and those that were developing investment proposals knew that it would be a tough competition to win. Just within the 'open' agenda, most institutions reported that the REF OA requirements currently dominate both operational capacity and leadership attention.
19. A considerable number of interviewees observed a change of PVC Research, causing a hiatus during the gap or a change of direction or emphasis. Whilst this did not necessarily prevent development or operational progress, it was reported to have affected institutional visibility, advocacy and academic engagement. Some PVCs were reported to be very engaged and actively involved, whereas others took a more distant stance, simply expecting RDM to be delivered. Only in a very small number of cases did there seem to be opposition at the leadership level to a policy of data sharing.
20. Library staff and their leaders tended to be much more engaged and supportive, which is not wholly surprising. Operational responsibility tends to be located in the Library, and occasionally in the Research Office or equivalent. In many institutions there is a strong relationship between the Library and the Research Office in delivering RDM, but in a reasonable number there was evidence of demarcation or disagreement. The role and level of involvement of the institution's IT Services also varied. These issues are discussed further in the RDM Services section below.

C.3 Academic Engagement with RDM

21. Academic engagement is key to delivery of RDM and its benefits. All institutions expressed variable levels and forms of academic engagement. In many cases there was 'conceptual interest, but not yet practical engagement', as one interviewee put it.
22. Part of the challenge is the current level of understanding of RDM, coupled with the elision of data management and data sharing. Researchers are reported to be relatively supportive of data management but more wary of data sharing. However, there is a perception that both are externally-applied requirements (i.e. 'more bureaucracy'), rather than being positive developments that can enhance their research and their reputation. Researchers therefore, for example, tend to create a Data Management Plan (DMP) where required by their funder, but not to use it actively thereafter, and not to create one where there is no funder (or institutional) mandate to do so.
23. This is possibly caused in part because advocacy for RDM is naturally led by those providing the operational support (e.g. Library staff), and less so by institutional leaders or senior researchers. Additionally, as noted in the previous section, meeting the OA requirements of the REF currently dominates both academic messaging and operational support (as OA and RDM is often supported by the same person or team).
24. Part of the variation in engagement reflects subject differences, but not necessarily as might be expected. The subject most frequently indicated as being least engaged with or most hostile to RDM was computing science. Arts and humanities researchers were often observed as needing to be helped to understand that they had data to be managed, but they were also reported as being interested to understand if and how RDM might apply to them. A small number of institutions were actively exploring what data and its management means in humanities subjects, as a joint exercise between the subject areas and the Library.

25. Some of the most innovative practices were being demonstrated by institutions with a significant creative practice presence: the ability to manage and represent data sets seems to provide better support for the outputs of creative research, such as portfolios, than do standard output repositories. Such institutions were therefore tending to use their data repository for outputs, rather than trying to use an output repository for data. They were also using the repository to support the research process and to reflect the creative process, and not just to act as a store for data to be curated.

C.4 Evidencing Benefits of RDM and Data Sharing

26. A number of institutions expressed the desire for more evidence of the benefits of RDM and data sharing, in order to convince researchers. Having the equivalent of the studies that show that OA improves citation rates were thought to be needed. A couple of individual cases were shared as exemplars, with a desire for more of these, across different subject areas. For example, Alasdair Rae from Sheffield², whose geospatial data has become widely used, and who was not initially enthusiastic about data sharing.
27. A small number of interviewees commented that some researchers had come to realise that they were finding it hard to reuse their own data a few years later, and hence that they needed to improve their data management (e.g. structure, cleaning and documentation) for their own benefits. This perhaps illustrates the need to find and share examples of benefits for the individual, rather than only expressing benefits in terms of wider society. Sharing with one's future self is clearly of self-interest, as well as being good practice. Other circumstances that might be used for exemplars include collaborations and partnerships that have only happened because of shared data.
28. In a small number of cases, an interviewee observed that researchers who depend on secondary data for their research tend to be more willing to share their own data. This may not be so surprising, and tends to be restricted to certain fields.
29. Interviewees from an IT background also provided examples of their advocacy in relation to protection against data loss or corruption, in particular through damaged hardware that has not been properly backed up.
30. As well as demonstrating the benefits to researchers of managing and sharing their data, and to others of that data being shared, it may also be necessary to demonstrate more clearly the benefits to institutions, e.g. in terms of the cost of academic time against that of good facilities and support, as well as the flexibility and capability (and competitive edge) that a well-managed data infrastructure and environment can provide.

C.5 Repositories for Data

31. The significant majority of interviewees had separate repositories for their outputs and for their data. In some cases these were different tools, but in many cases it was two instances of the same tool. In a small number of cases, the two instances were hosted by different providers.
32. The decision to have separate repositories seems to have come about for a number of related reasons. Several interviewees expressed a view that the two sets of information are different and should not be mixed. This was often coupled with a concern about data deposits corrupting the output deposits, which needed to be carefully controlled and managed in the context of REF. There were also views that the metadata for data deposits needed to be different from that for outputs, i.e. could be a lot simpler. This

² https://www.sheffield.ac.uk/usp/staff/alasdair_rae

was often expressed in the context of academic dislike of the institution's output repository and its deposit process. Quite a few of those initially supporting separate repositories also noted that if they were starting now, they might well choose to have a single repository. This seemed to be a combination of experience gained and a perception of an improvement in some repository tools. Indeed, a large proportion of interviewees expressed some level of concern about their output repository tool, which strongly influenced their decision about their data repository. Based on the interviews (not on a detailed market survey), the historical market for data management tools and repositories does not seem to be that mature, both in terms of customer requirements and supplier provision.

33. Those Libraries with their own systems / development capacity tended to be more comfortable with their repository, because they felt more in control, and did not have to rely on their institutional IT service. Smaller institutions noted the need for externally-hosted services as their institution does not have the capacity to provide and support one.
34. A different driver for separate repositories applies in a relatively small number of institutions, where different offices are responsible respectively for outputs and for data. Where there is this separation, it is usually the Library who is responsible for outputs and the Research Office for data. Whilst there may be strategic or operational logic for this, it was often also indicative of a fragmented approach to the Open agenda.
35. Plan S was mentioned on a relatively small number of occasions (note that its requirements developed during the course of the interview process). Some interviewees felt that it would disrupt the provision of output repositories, causing a considerable number of institutions to change their repository. In doing so, institutions, in their view, would be looking for a single repository for outputs and data. Their caveat was that repositories would not be changed until after the current REF. Conversely, a couple of interviewees felt that it would not affect data repositories to any great extent.
36. A number of institutions with strong specialisms in creative and arts subjects observed that they had found data repositories helpful in storing and profiling their outputs, because of their better ability to handle portfolios and similar. Some of the interesting, and creative, use of data repositories was happening in these institutions, who had typically been involved in the Kultur³ and Kaptur⁴ projects previously supported by Jisc.

C.6 RDM Services

37. All interviewees spoke of providing a range of services in relation to RDM. These typically included policy, advocacy, training, data management plan (DMP) support, deposit, and curation. Most interviewees noted the central role of human interactions, as opposed to technical solutions, in the delivery of these services.
38. In the majority of cases, these services were provided by the Library, typically by the person or team who were also responsible for OA (whether entitled scholarly communications or not). The exception tended to be in relation to policy, which was often the responsibility of the Research Office. As already noted, in a minority of cases RDM was being led from the Research Office because of the funder requirement for it. Research Office involvement included variously those with responsibility for policy, REF, research governance, grants, and business engagement. The specific focus of the individual(s) concerned made a difference as to their attitude and approach to RDM.

³ <http://kultur.eprints.org/>

⁴ <https://vads.ac.uk/kaptur/about.html>

Someone responsible for OA or the REF may be attuned to the societal benefits of data sharing whilst a person concerned with data security or one driven by commercial outcomes are likely to have different focal points.

39. There was often, but not always, a relationship between the Library and the Research Office to ensure that grant applicants received RDM support (e.g. in relation to their DMP). Many institutions made use of DMPonline, with a split between those who subscribed and those who did not. Very few interviewees spoke of making institutional use of the DMPs to help them manage their deposits and to maximise the use and hence value of their data sets. This was not explored, but is likely to be because of time pressures and because of a compliance-based approach. One interviewee observed that DMPs are more likely to be maintained and used by large projects with a dedicated project manager.
40. All institutions accepted deposit in an appropriate subject repository or equivalent. One had started with an intention to mandate deposit of all data sets in the institutional repository, but had then relaxed their approach. Interestingly, not all mandated the deposit of the meta-data in the institutional repository, although they generally expressed that they would like it to be deposited. This potentially reflects the state of play in many institutions, in trying to be as encouraging of deposit as possible, rather than over-emphasising full compliance. It does mean, though, that quite a few institutions do not have a complete register of the data sets produced by their researchers.
41. The operational team(s) were generally responsible for advocacy and training, with PVCs engaged in a few cases. Whilst training was provided, many interviewees observed the difficulty in getting established researchers to attend, and that courses were typically targeted at PGR students and early career researchers.
42. Very few interviewees spoke of, or had considered, services and tools aimed at data capture or analysis. This reflects the earlier observation that institutional approaches have tended to be focused on policy and deposit in a repository. However, it is these earlier tasks that are of more interest (and practical benefit) to researchers. One interviewee noted that they had responded to researchers' needs by enabling the sharing of active data with collaborators rather than taking a rigid approach to deposit of final data. It might be of greater value to institutions, funders, policy-makers and suppliers to focus a little more attention on this point of the data lifecycle.
43. The role of the IT service varied, but typically focused on the provision of active data storage and support of the relevant software (either directly or in terms of managing the external supplier). In a minority of cases there was clear evidence of a close and productive working relationship between the IT service and the team(s) responsible for RDM. Conversely, a number of interviewees expressed concern about their IT services' approach, such as only being interested in the physical or virtual hardware required and not understanding the policy context. For example, in being concerned that data that was stored was not being used (hence 'wasting space'), irrespective of the funder or institutional requirement to maintain the data set for a considerable period. A different example was of the implementation project being led (at the requirement of the PVC) by the IT service, which exacerbated the technical solution approach to the detriment of addressing the cultural and behavioural issues.
44. Two seemingly opposing, but actually complementary, views were expressed by two institutions. One suggested that IT would only take a system / service seriously if it was deemed to be 'core' and paid for from their budget. The other felt that IT could not

support all necessary systems as 'core', which is why externally-hosted systems were a positive way forward.

45. Discussion about the provision of active data storage raised a number of interesting points. Institutions were divided between those who automatically provided space to researchers and those that required a case. In only a few instances, the process to release space was controlled by the Library; generally it was in the IT service's purview.
46. Where space was readily available and expanded to meet the needs of the research, interviewees tended to express concerns about convincing researchers to migrate their data from short-term storage to long-term curation, as there was no real pressure or incentive to do so. In a small number of cases, the availability of space was driven by external capital funding provision for data storage, rather than by strategic design.
47. Returning to the earlier comments about location of responsibility, one can observe that RDM requires or touches on a range of expertise and areas of a university's business. Rather than expecting all aspects to be covered by one service, institutions need to be better at integrating the viewpoints and expertise, both to manage their research data better and to gain the most benefits from it. Each area of competence needs to be involved in an informed dialogue that crosses the boundaries of the organisational structures; for example:
 - IT on physical and virtual infrastructure;
 - Library on information management, OA and long-term curation;
 - Research support on funding policy and applications;
 - Business development on external relationships and use;
 - Research governance on information management compliance.

C.7 Digital Preservation

48. Of all the areas discussed, digital preservation was the least well developed, as acknowledged by the interviewees. A considerable proportion of institutions have not got any digital preservation capacity, activities or plans in place. A small number have some functionality, but do not have a broader or longer-term plan for curation and preservation. Those institutions with more developed capacity or plans tend to be those with either a Special Collections function or with an active institutional Archive function, or both.
49. All institutions who were investigating digital preservation, including those who might be described as being the more advanced of those interviewed, said that they were still at the exploration stage. One institution, for example, commented that they had chosen to develop the skills of their staff before investigating the specifics of any tools, thus making them better able to develop their institutional approach and to assess the available technology.

C.8 Functionality Required

50. The expression of need for functionality over and above the basics of a repository centred around integration: with the output repository, the research information systems, the broader research management system, as well as the HR system and sometimes the Finance system.
51. External integration was also relevant, given that all of the interviewed institutions accept deposits in subject repositories. Automatic deposit of the metadata into institutional repositories would therefore be welcomed.

52. Increased automation, leading to cost savings, was frequently expressed. This seems to reflect the currently manual and fragmented nature of much of the process. The need for DOI minting was mentioned by some interviewees, but others noted that they had already subscribed to a service.
53. One institution expressed (and some others implied) the need for an 'actual technical solution', rather than just policy positions, for example in relation to (active) storage that enables sharing outside the institution (i.e. with collaborators). This was the most common comment relating to support of the earlier stages of the data lifecycle. A small number of interviewees also expressed the desire for tools to support data capture, integration and analysis.
54. A number of institutions have NHS Information Governance-compliant data storage, and how this needs to sit alongside the broader data storage typically needs some careful thought.

D. Conclusions

55. The process of interviewing 47 institutions has been interesting and enlightening. Progress is being made, but slowly and often in constrained circumstances. Most institutions are working towards compliance rather than seeking to enable research and its possible outcomes.
56. Few institutions selected for interview are considering all elements across the data lifecycle, typically only really addressing two of the six segments shown in Figure 1, and approaches are too often fragmented. The sector is currently scratching the surface of the full spectrum of functionality and hence benefits across the data lifecycle.
57. The challenges are both technical and cultural, with appropriate academic behavioural motivations not being in place. Funders, policy-makers and institutional leaders have not adequately recognised the nuanced drivers that are required to get positive engagement and full benefit. One interviewee's message to funders, wishing to be quoted, was: "For goodness sake will you standardise".
58. The compliance-focused approach has led some institutions to believe that having a policy and a repository is sufficient. That does not equate to good research data management or to shared, open data. Whilst the two (RDM and open data) are related, there has been too much conflation of the topics and of their description and advocacy. Sharing data creates different academic concerns (e.g. inappropriate use, loss of leadership) compared to that of RDM (e.g. perceived as a bureaucratic burden), and hence each needs to be addressed separately.
59. There is a relatively small, sometimes isolated or fragmented capacity to support active RDM and data sharing. Institutions need to develop a fuller, more holistic understanding of their research (and other) data, and to draw on the full range of expertise in an integrated manner to manage their data and to gain the most benefits from it: IT on physical and virtual infrastructure; Library on information management, OA and long-term curation; research support on funding policy and applications; business development on external relationships and use; research governance on information management compliance.
60. Institutions, funders, policy-makers and suppliers should focus more attention on researchers' needs, in particular around active data capture, integration, management and sharing. Addressing this point in the lifecycle is more likely to gain academic

engagement, and also enhance the quality of the data, thus making it more worth sharing.

61. Institutions might also make more use themselves of the data management plans, to manage their deposits and to maximise the use and hence value of their data sets.
62. The potential for research data management and sharing exists but is untapped because it is inadequately recognised and supported. There are undoubtedly peaks of good practice, but the baseline needs to be lifted. It behoves those responsible for the research ecosystem to address the substantive issues in order to unlock this potential.

E. Appendix: Institutions Interviewed

Aberystwyth University	University of Dundee
Anglia Ruskin University	University of East Anglia
Bangor University	University of Exeter
Bath Spa University	University of Greenwich
Cardiff Metropolitan University	University of Hertfordshire
City University	University of Keele
De Montfort University	University of Kent
Edge Hill University	University of Leeds
Edinburgh Napier University	University of Leicester
Glasgow Caledonian University	University of Liverpool
Glasgow School of Art	University of Newcastle-upon-Tyne
Kingston University	University of Northumbria at Newcastle
Liverpool John Moores University	University of Oxford
Liverpool School of Tropical Medicine	University of Portsmouth
London School of Hygiene and Tropical Medicine	University of Reading
Loughborough University	University of South Wales
Open University	University of Southampton
Oxford Brookes University	University of Stirling
Queen Mary University of London	University of Strathclyde
Royal Veterinary College	University of Sussex
Sheffield Hallam University	University of the West of England, Bristol
University of Bedfordshire	University of Ulster
University of Birmingham	University of Warwick
University of Brighton	

Dr Ian Carter
 Director, Carter Research Navigation Ltd
 28/2/20