The International Journal of Digital Curation
Issue 2, Volume 6 | 2011

Making Sense: Talking Data Management with Researchers

Catharine Ward & Lesley Freiman,

University of Cambridge

Sarah Jones, Laura Molloy & Kellie Snow,

University of Glasgow

Abstract

Incremental is one of eight projects in the JISC Managing Research Data programme funded to identify institutional requirements for digital research data management and pilot relevant infrastructure. Our findings concur with those of other Managing Research Data projects, as well as with several previous studies. We found that many researchers: (i) organise their data in an *ad hoc* fashion, posing difficulties with retrieval and re-use; (ii) store their data on all kinds of media without always considering security and back-up; (iii) are positive about data sharing in principle though reluctant in practice; (iv) believe back-up is equivalent to preservation.

The key difference between our approach and that of other Managing Research Data projects is the type of infrastructure we are piloting. While the majority of these projects focus on developing technical solutions, we are focusing on the need for 'soft' infrastructure, such as one-to-one tailored support, training, and easy-to-find, concise guidance that breaks down some of the barriers information professionals have unintentionally built with their use of specialist terminology.

We are employing a bottom-up approach as we feel that to support the step-by-step development of sound research data management practices, you must first understand researchers' needs and perspectives. Over the life of the project, *Incremental* staff will act as mediators, assisting researchers and local support staff to understand the data management requirements within which they are expect to work, and will determine how these can be addressed within research workflows and the existing technical infrastructure.

Our primary goal is to build data management capacity within the Universities of Cambridge and Glasgow by raising awareness of basic principles so everyone can manage their data to a certain extent. We will ensure our lessons can be picked up and used by other institutions. Our affiliation with the Digital Curation Centre and Digital Preservation Coalition will assist in this and all outputs will be released under a Creative Commons licence.[1]

The *Incremental* project

Incremental is a collaboration between Cambridge University Library and the Humanities Advanced Technology and Information Institute at the University of Glasgow.[2,3,4] As the project name suggests, we are adopting a step-by-step approach to research data management, proposing that top-down, policy-driven, or centralised solutions are unlikely to prove as effective as clear, appropriate and practical support delivered to researchers in a timely manner.

The initial phase of the project focused on gathering researchers' requirements. A digital preservation scoping study (Jones, 2009) was undertaken at the University of Glasgow in 2009 with support from its Digital Preservation Advisory Board. This work fed into the *Incremental* project and helped to guide a comparable scoping study at the University of Cambridge.[5] A total of 25 semi-structured interviews were completed at the University of Glasgow and 29 at the University of Cambridge. We spoke with a mixture of Ph.D. students, early career and senior researchers, and departmental computing officers, mostly in one-to-one interviews but occasionally in groups. Similar research groups and departments were chosen in the two studies to allow for comparisons across disciplines as well as between institutions.[6]

In order to understand researchers' current practices, concerns, and needs for data management, we asked what type and volume of data they create, how these are organised, stored and backed-up, and their plans for data sharing and preservation. We also investigated researchers' awareness of existing support and guidance, and asked for their views on its usefulness. We have followed up preliminary findings in this area by undertaking a review of the data management guidance offered by Russell Group universities (a group of twenty leading UK universities "committed to maintaining the very best research") and by running observations to identify how and where researchers look for support.[7]

The current phase of the project (the implementation phase) responds to the needs articulated by researchers. We have identified and assessed existing guidance, training, and support at each of the two institutions. A key goal of the implementation phase is to raise awareness of and access to existing support by matching it more closely to user needs where possible, rather than creating new content.

We have uncovered a lot of useful guidance from support services at both of our institutions but this is typically hard to find and tends to be uncoordinated, scattered across IT services, libraries, research support, offices which advise on data protection and Freedom of Information (FoI) legislation, and individual departments. We have re-positioned — and, where necessary, added to — these distributed and fragmented resources to form a comprehensive and coherent set of guidance and links at each institution. We have also sourced and repurposed external resources to fill gaps where required.

---

[2] Incremental: http://www.lib.cam.ac.uk/preservation/incremental/cambridge.html.
[3] Cambridge University Library: http://www.lib.cam.ac.uk/.
[4] Humanities Advanced Technology and Information Institute: http://www.gla.ac.uk/departments/hatii/.
[5] This work has build on lessons from the Data Asset (formerly Audit) Framework. See specifically the DAF Implementation Guide available at: http://www.data-audit.eu/documents.html.
[6] These included Archaeology, English, Chemistry, Engineering, Public Health, Ecology & Evolutionary Biology and Environmental Science.
[7] Russell Group: http://www.russellgroup.ac.uk/home.

We plan to develop additional resources through our training and workshop programmes. Our first training course will take place in November 2010 at the University of Glasgow and a series of thematic and discipline-specific seminars covering topics such as FoI, copyright, and dealing with sensitive data are being planned for spring 2011.[8]

What We Found

The findings of our studies echo the conclusions of previous work, such as the *Investigating Data Management Practices in Australian Universities* study (Henty, Weaver, Bradbury, & Porter, 2008), Data Audit Framework pilots (Jones, Ball, & Ekmekcioglu, 2008), and sharing, curation, re-use, and preservation discipline-specific case studies.[9] We found that many researchers organised their data in an *ad hoc* fashion, simply doing whatever seemed easiest at the time. A lack of clear file naming practices and version control often led to difficulties later on when researchers tried to find and re-use their own legacy data. Many also felt that there was not enough institutional storage and sought cheap solutions, ranging from computer hard drives, laptops and external hard drives to memory sticks and free email accounts – often without realising how this put their data at risk in terms of security and back-up.

Considering the longer-term potential for re-use of data, most researchers were positive about data sharing in principle, though almost universally reluctant in practice. Most perceived data sharing as an activity requiring careful preparation, annotation, and contextualisation of data for little professional recognition. Publication of peer-reviewed papers was seen as the primary way of gaining prestige in most disciplines.

The misconception that back-up is equivalent to long-term preservation was almost universal. Researchers were often uncertain about which digital formats are better for preserving data and found the need for metadata and documentation to be a formidable barrier to depositing in data centres and repositories. We intend to focus on the benefits that good documentation brings by enabling researchers to effectively re-use *their own* data and highlight re-use by others as a secondary benefit.

On the whole we found researchers' needs to be surprisingly common across disciplinary and institutional boundaries. Two key themes emerged from the requirements-gathering phase and are expanded on below.

Language Matters

It became apparent in conversations with researchers that most do not understand what 'digital curation' is, nor are they familiar with terms such as *preservation* or *digital repository*. Use of the term *research data* has also proved problematic in some cases, particularly when speaking with humanities researchers who tend not to think of the digital documents they produce as 'data'. When discussing possible ways to support researchers, many seemed suspicious of 'policies', which imply a mandate, but were more receptive to 'guidance' or 'advice' which may be essentially the same thing but convey a sense of purpose and assistance rather than requirement. It became clear, therefore, that we should translate research data management vocabulary from the specialist to non-specialist and adopt a more open, active tone in guidance resources.

---

[8] Digital Preservation Training Course: http://www.dptp.org/.
[9] A series of seven immersive case studies to identify disciplinary approaches to data deposit, sharing and re-use, curation and preservation. Available at: http://www.dcc.ac.uk/projects/scarp.

Subsequent research has uncovered multiple uses of the term *data management*.[10] In several UK universities, it is used to mean how a researcher goes about gathering, recording, ordering, and manipulating their information, as they work on their project. Much of this could arguably be termed *research methods*. In other institutions, the term is used to mean how one looks after the data after the work is completed – how one takes care of its longevity, access, integrity and security. Some other institutions include both these processes to support a third, inclusive view, that is, that data exists within a lifecycle, beginning with the creation or receipt of the data, and finishing with how to look after it in perpetuity (or until it should be disposed of). These are three distinct uses of the one phrase. If we are to support researchers to manage their data, more clarity is needed on the processes, roles and responsibilities that data management encompasses.

The Need to Start Early – in the Research Lifecycle and in Researchers' Careers

In order for data management to be most effective, researchers and research support staff should plan for it from the outset. Consent agreements and decisions about how to create data in the short-term affect what is feasible with respect to its re-use in the longer-term, and so need to be approached with an appreciation of future data sharing and preservation planning. Many research funders in the UK have responded to this need by instituting data management and sharing plan requirements. Bodies such as the Digital Curation Centre and UK Data Archive are playing a key advocacy role to raise awareness of these requirements and support researchers to create and implement data management plans.[11,12] *Incremental* has created complementary guidance to help researchers at our institutions understand and navigate this landscape.

The majority of researchers who participated in our studies at the Universities of Cambridge and Glasgow suggested that the best possible point at which to to intervene with guidance and training is very early on in a researcher's career. This was also confirmed during interviews with those providing support and advice to researchers. Early-career researchers tend to shoulder the responsibility for the day-to-day management of research data and this is the point at which habits begin to form. Some commented that more senior researchers' practices are fully entrenched and could therefore be difficult to influence, but they expected that senior researchers would support efforts to improve practices in the new generation of researchers.

What Researchers Want

Help Researchers Can Find and Understand - When They Need It

Many researchers raised concerns that it is difficult to find relevant guidance when they need it. At the Universities of Cambridge and Glasgow, data management guidance is often scattered across different internal and external websites. This situation appears to be common for universities across the United Kingdom. We have studied the data management support openly available on the websites of the Russell Group universities, and found that only two provide clearly-signposted research data management support resources on their institutional websites.

---

[10] This is discussed in an *Incremental* blog post at:
http://incrementalproject.wordpress.com/2010/07/14/vocabularyjargonterminology-synonyms-and-specialist-language/.
[11] Digital Curation Centre: http://www.dcc.ac.uk/.
[12] UK Data Archive: http://www.data-archive.ac.uk/.

Many researchers from our scoping study requested brief, practical advice in formats that make it easy to access relevant information quickly, as they do not have time to read through long policy documents. Several found it difficult to obtain a simple, unambiguous answer to their queries once they had located relevant guidance. They suggested a variety of alternative formats such as factsheets, Frequently Asked Questions, checklists, crib sheets, flow diagrams, bulletins, newsfeeds and email alerts. Rather than creating new resources, we have summarised and linked to existing guidance, adding to and enhancing where required so it is easier to find and more user-friendly for researchers. The guidance points primarily to institutional support, as a key aim is to raise awareness of this, but also includes links to particularly useful external resources.

Researchers were also keen for diverse, web-based modes of training such as online tutorials, videos and interactive learning resources. One researcher commented:

> "There's no point being told all this stuff when you're not using it
> because – I mean for me, it goes in one ear and out the other. I
> only learn how to do things when I need to know".
> (Incremental, 2010)

The timeliness of training is paramount. To ensure researchers can access training as and when required, online training materials will be developed from the Digital Curation Centre-supported courses we run.[13,14] We are also investigating online learning resources that can be developed as part of our seminar series. To maintain continuity for researchers and make our training resources sustainable, *Incremental* is also adopting a 'train the trainer' approach, working with existing training providers to create basic slides and resources targeted at researchers. These will be designed to be dropped into a variety of training courses, incorporating data skills at relevant points in research training.

Tailored Support to Help Researchers Make Informed Choices

Several researchers requested tailored support. One interviewee explained that no matter how clear the guidance, people will always want to pick up the phone and ask someone how it relates to them. To address this practice we have ensured each of our guidance webpages ends with the contact details of University support staff who can help with the particular issue at hand. Many interviewees also wanted a university body or support network to which researchers could be referred.

We are encouraging researchers to use tailored support services, as the context of each case leads to different advice. Indeed, support staff at the University of Glasgow confirmed it is crucial to understand the needs of each individual project, as it is only by understanding the skills, resources and needs of the research team together with their project aims that appropriate recommendations can be given.

---

[13] Managing your Digital Research Data (DC101 Lite):
http://www.dcc.ac.uk/events/workshops/managing-your-digital-research-data.
[14] Managing your Research Data: Digital Curation:
http://www.dcc.ac.uk/events/workshops/managing-your-research-data-digital-curation.

How We Are Responding

We are currently in our implementation phase, which has three main objectives:

1. re-position existing guidance so researchers can find the advice they need;
2. connect researchers with one-to-one advice, support and partnering;
3. offer practical training and seminars with discipline-specific examples.

The focus throughout these activities will be to build on the lessons of earlier projects and enhance existing support and guidance rather than starting afresh.

Re-position Existing Guidance so Researchers Can Find the Advice They Need

In our initial scoping work, we repeatedly found that researchers were unaware of much of the support, guidance and services that are currently available to them through their institutions. This was affirmed through observations in which we asked researchers to perform a number of simple information-seeking tasks, related to data management, to see if they could find relevant information through the University website. Researchers were often confused about where to look, and when they found resources, reported that guidance was quite difficult to skim through in order to find the salient points.

We chose to create a coherent set of data management support pages at each institution to address these concerns. We looked at models from within the data curation sector to inform this work and reviewed commercial websites such as Google and Flickr for their clean design and navigation options.[15,16] The data management support pages are organised into four themes:

- Creating your data: data planning, file formats, intellectual property rights, ethics, data protection, and FoI;
- Organising your data: file naming, organising data, and documentation;
- Accessing your data: storage, remote access, and security;
- Looking after your data: back-up, selection, preservation, and data sharing.

In keeping with researchers' wishes for simple guidance, the web pages contain clear and concise information. We have kept digital curation terminology to a minimum. Where useful, we provide multiple terms or an explanation of essential specialised terms for example *metadata*, a term which is familiar to many researchers in the sciences, but not the humanities. We have also chosen to employ the phrase 'looking after your data' rather than 'preservation', recognising that preservation approaches are not very formalised at present, that many researchers are unfamiliar with the term *preservation* in a data context, and that few have thought about 'preserving' their data for the long-term.

---

[15] See for example the Massachusetts Institute of Technology data management and publishing webpages: http://libraries.mit.edu/guides/subjects/data-management/ and Edinburgh University guidance:
    http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt
[16] Google: www.google.com; Flickr: http://www.flickr.com/.

Individual resource pages for each topic employ a Frequently Asked Questions format to make it easier to access relevant information at a glance. There are links to further internal and external guidance for those who want more detailed or nuanced information, and helpful support contacts within the institution. Additional links in the sidebar of the page provide researchers with contact information and training resources. We are also considering a printed version that could be provided on training courses and in the induction packs for new researchers at our universities.

The resources which we are creating are modular and will be released under Creative Commons licences to encourage others to repurpose and adapt them within their own institutions.[17] We also wish to encourage staff within our institutions to take our resources and embed them within local pages and training programmes. At the University of Cambridge, the Research Data Management Training Materials project, DataTrain, will be able to take this localisation forward beyond *Incremental* within the University's social anthropology and archaeology departments.[18] The resources created will also be closely associated with those of the institutional repository service (Dspace@Cambridge), to take advantage of its profile and advisory services and to ensure sustainability beyond the project.[19]

Connect Researchers With One-to-One Advice, Support and Partnering

During interviews and in the website observations, researchers have explained that they would look for somebody to call as they find it more reassuring to present a case and get an authoritative answer on those particular circumstances than trying to read general guidance and extrapolate what is relevant. We have responded to this by making the links to helpdesks and contact details for named support staff much more prominent.

We have also focused on raising awareness of tailored support services, such as the Resource Development Officers at the University of Glasgow. By working with College research offices within each institution we can make sure that researchers are directed to existing support staff for one-to-one advice during the proposal-writing stage of projects and beyond. This will include the provision of training for research office staff to ensure that they are aware of relevant funding body policies and can provide basic advice at the grant application stage.

Research support staff reported that providing assistance in this way can enhance how research is conducted, as researchers can be informed of relevant technologies and tools of which they were unaware. Encouragement for planning data management from the outset of projects will also allow researchers and support staff to budget for relevant technical support, storage and data preparation costs.

Offer Practical Data Training with Real-Life, Discipline-Specific Examples

Researchers are keen for pragmatic, best-practice guidelines, and flexible modes of training. We feel that a mixture of online training resources and face-to-face sessions would be benefical. We are therefore producing factsheets, online screencasts and case studies to provide a suite of online training resources that will be accessible when researchers need them, as well as running training courses, and a series of data management-related seminars.

---

[17] Creative Commons: http://creativecommons.org/.
[18] DataTrain: http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain/datatrain.aspx.
[19] Dsapce@Cambridge: http://www.dspace.cam.ac.uk/.

For the hands-on training, we aim to embed data training early in the research lifecycle and to target PhD students and early-career researchers, as our scoping work has indicated that (a) these groups are often given responsibility for data management, and (b) their habits are still forming, so they may be more receptive to guidance than established researchers. A course on curation for researchers is being run at both institutions to inform researchers and those who provide support about the issues of data management.

We are also planning a number of seminars for spring 2011. Thematic seminars will be held at the Centre for Arts, Social Sciences and Humanities in the University of Cambridge, addressing issues such as intellectual property rights, FoI, ethics, and sensitive data, while at the University of Glasgow, the seminars will bring researchers together along disciplinary lines so subject-specific norms and best practice can be explored. We hope to collaborate with the JISC Research Data Management Training projects on this as they plan to develop subject-specific training materials.[20] Framing data management within particular themes and disciplines is hoped to appeal to researchers more than generic data management workshops.

All seminars will provide case study presentations to give practical examples of how researchers are addressing their data management concerns. We will work with the research groups involved in our project to provide these, and where appropriate, bring in external speakers from the Managing Research Data programme or beyond to provide inspiring solutions that address concerns we have encountered. We plan to create online resources from the seminars such as video interviews, case studies, and demonstrations to ensure continued engagement and benefit.

Conclusions

The requirements-gathering work confirmed that there is a great need for support, and that researchers want basic assistance to help with the day-to-day issues faced when creating and managing their data. Addressing these issues need not be overly expensive or depend upon the implementation of highly technical solutions; indeed, our findings suggest that data management is in many respects a 'people problem' rather than a 'technical problem'. We believe collating and repurposing existing guidance, training, and support so that it is easy to access, clear, engaging, and relevant to researchers will be effective in the long run.

Acknowledgements

---

[20] Research Data Management Training Materials: http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain.aspx.
[21] Digital Preservation Coalition: http://www.dpconline.org/.

References

Henty, M., Weaver, B., Bradbury, S., & Porter, S. (2008). *Investigating data management practices in Australian universities.* Report of Australian Partnership for Sustainable Repositories Online Research Collections Australia Support Network. Retrieved July 21, 2010, from http://www.apsr.edu.au/orca/investigating_data_management.pdf.

Incremental (2010). *Scoping study report and implementation plan.* Retrieved July 21, 2010, from http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_170910.pdf.

Jones, S., Ball, A., & Ekmekcioglu, C. (2008). The data audit framework: a first step in the data management challenge. *International Journal of Digital Curation, 3*(2), 112-120. Retrieved July 21, 2010, from http://www.ijdc.net/index.php/ijdc/article/viewFile/91/62.

Jones, S. (2009). *Findings from scoping study investigations to inform the development of a digital preservation strategy at the University of Glasgow.* Unpublished manuscript, University of Glasgow, UK.