



University
of Glasgow

Davidson, J. (2013) *Supporting emerging researchers in data management and curation*. In: Mackenzie, A. and Martin, L. (eds.) *Mastering Digital Librarianship: Strategy, Networking and Discovery in Academic Libraries*. Facet, London. ISBN 9781856049436

Copyright © 2013 Facet Publishing

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

Content must not be changed in any way or reproduced in any format or medium without the formal permission of the copyright holder(s)

When referring to this work, full bibliographic details must be given

<http://eprints.gla.ac.uk/80150/>

Deposited on: 22 October 2014

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Supporting emerging researchers in data management and curation

Joy Davidson, Digital Curation Centre (DCC)

While scholarly publishing remains the key means for determining researchers' impact, international funding body requirements and government recommendations relating to research data management (RDM), sharing and preservation mean that the underlying research data are becoming increasingly valuable in their own right. This is true not only for researchers in the sciences but also in the humanities and creative arts as well. The ability to exploit their own - and others' - data is emerging as a crucial skill for researchers across all disciplines. However, despite Generation Y researchers being 'highly competent and ubiquitous users of information technologies generally' they appear to be a widespread lack of understanding and uncertainty about open access and self-archived resources (Jisc study, 2012). This chapter will consider the potential support that academic librarians might provide to support Generation Y researchers in this shifting research data landscape and examine the role of the library as part of institutional infrastructure.

The changing landscape will impact research libraries most keenly over the next few years as they work to develop infrastructure and support systems to identify and maintain access to a diverse array of research data outputs. However, the data that are being produced through research are no different to those being produced by artists, politicians and the general public. In this respect, all libraries - whether they be academic, national, or local - will need to be gearing up to ensure they are able to accept and provide access to an ever increasing range of complex digital objects.

First, librarians need to understand what research data might include

A shared concept of what research data are is essential for ensuring that researchers are able to adhere to institutional and funding body policies and mandates. However, determining exactly what is meant by the term research data and what sorts of research outputs need to be considered is an ongoing challenge. Definitions vary across disciplines, funders, and publishers. For anyone working to support researchers in managing and sharing their data within a Higher Education Institution (HEI), it is essential that a clearly defined view of what constitutes research data is agreed early and endorsed at the right levels (Research Group, School, College, University-wide).

As a general guide, the Digital Curation Centre (DCC) recommends that researchers should consider how they will maintain access to any research data that may be necessary for enabling the validation of their published research findings. At the University of Glasgow, research data are defined as: 'any material (digital or physical) required to underpin research. For different disciplines this may include raw data captured from instruments, derived data, documents, spreadsheets & databases, lab notebooks, visualisations, models, software, images, measurements and numbers' (University of Glasgow Draft Research Data Policy, 2012).

While the term research data is something that the majority of STEM researchers are comfortable with as a description of their academic outputs, many researchers in the arts don't necessarily see themselves as data producers. However, there are ongoing efforts to better define research data for arts researchers. For example, the Jisc-funded KAPTUR project is exploring the notion of research data management for the performing arts. The project is being led by the Visual Arts Data Service and project partners include the Glasgow School of Art, the University of the Arts London, the University for the Creative Arts and Goldsmiths, University of London. Project partners have worked to develop research data management policies that include definitions of what might constitute research data. For instance, the University of the Arts London policy states that non-digital materials such as sketch books are covered in their Research Data Management Policy. Research being

undertaken by DCC staff in collaboration with the KAPTUR project has also helped to illuminate other possible definitions for research data outputs within the arts.

“documenting the research process’ & ‘visualisation and documentation’ were offered as alternatives to research data... Idea of ‘organisational moments’ or ‘trigger points for data creation or management activity.’” (Guy et al, 2013)’

Whatever your institution decides to include in its definition of research data outputs, it will be vital that all institutional support staff are aware of these definitions and can explain them to researchers as and when needed.

Next, librarians need to understand the range of research data management and curation activities

Once institutions have reached agreement on what they considered to be research data, the next step is to identify what support will be needed to ensure that these data are actively managed and curated as required by funder and publisher mandates as well as institutional policies. The DCC defines curation as ‘the active management and appraisal of data over the lifecycle of scholarly and scientific interest’ (DCC, n.d.). Data management and curation involves many stakeholders including IT staff, library staff, research administrators, senior managers to name but a few. It will be crucial that these varied stakeholders communicate effectively with each other and understand their roles and responsibilities in supporting research data management and sharing. The library is generally well placed to serve as a liaison point between the various stakeholder groups within an institution. The DCC developed the Curation Lifecycle Model (Higgins, 2008) to help break down the wide range of curation activities so that specific roles and responsibilities can be identified and assigned. The model has data at its core and, as such, can assist institutions to articulate and refine research workflows and to develop support services that enable the wide range of stakeholders involved to meet data management and curation requirements.

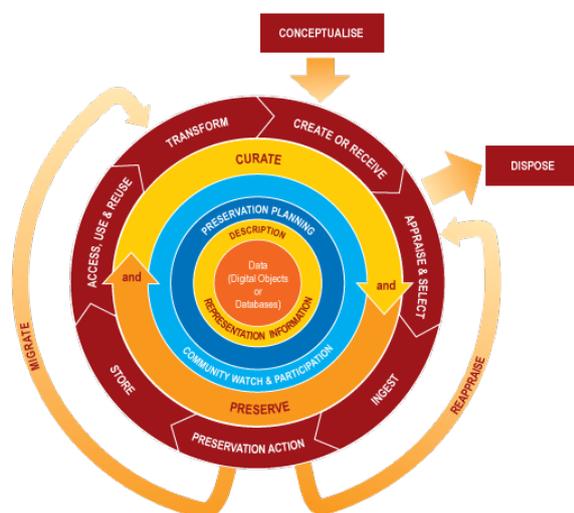


Figure 1, DCC's Curation Lifecycle Model

The model focuses on nine key areas of activity ranging from conceiving new research projects through to providing longer term access and facilitating reuse. These actions are depicted in the outer ring of the model (see figure 1). The model is cyclical in nature and it should be noted that users may engage with the model at any stage, not necessarily at the outset of new research activity - the conceptualisation stage - which is the notional ideal. In reality, there will likely be many cases where management and curation are not considered until the point of ingest into a repository and repository staff will then need to work back through the model to identify additional contextual information needed to make the data

accessible and reusable. Another point to bear in mind is that, as data-driven science becomes more commonplace, researchers may begin the curation lifecycle at the 'transform' stage by accessing and reusing others' data to lead to new research questions.

Librarians have an obvious role to play in ingesting, storing and providing access to research data that is deposited by researchers. However, by seeking to become more involved during the grant application and active research phases, librarians could help researchers to plan for increased access and use of their research outputs by making them more visible and understandable to external users. A more in-depth overview of the Curation Lifecycle Model is provided by Sarah Higgins in chapter two of the Facet publication *Managing Research Data* (Pryor, 2012).

Librarians should understand the range of factors driving the need for research data management and curation and be able to communicate these to researchers

In recent years, research data management and sharing have emerged as key themes in a number of countries worldwide. In 2007, the Organisation for Economic Co-operation and Development (OECD) declared that it viewed research data as a public good (OECD, 2007). Since then, a number of funding bodies in the US and UK have mandated that data management plans be submitted at the grant proposal stage to ensure that data generated through publicly funded research will be accessible and reusable over time – usually a period of at least 10 years. In the US, the National Science Foundation (NSF) and National Institutes for Health (NIH) have been mandating data management plans be submitted with grant applications since 2009. From 2013, the NSF will also require:

'For all new grant applications from 14 January, the US National Science Foundation (NSF) asks a principal investigator to list his or her research "products" rather than "publications" in the biographical sketch section. This means that, according to the NSF, a scientist's worth is not dependent solely on publications. Data sets, software and other non-traditional research products will count too.' (Piwowar, 2013)

The inclusion of research products as part of a researcher's biographical sketch is an encouraging development and hopefully will be an approach that is replicated by UK funders in the near future. In Europe, Horizon 2020 has made open access to research outputs a key objective. Indeed, data management plans will likely be required as part of funding applications in the European Commission's FP8 programme. A number of high profile reports have examined the need for improved infrastructure and research data management skills in the European landscape (Riding the Wave and Surfboard for Riding the Wave reports). Australia has invested heavily in developing and progressing a national approach to research data management through the establishment of the Australian National Data Service (ANDS). Canada established Research Data Canada in 2011 to begin to formalise a national approach to supporting research data management and sharing. However, some of the most dramatic advances in the past two years have been seen in the UK. In 2010, Research Councils UK (RCUK) issued their Common Principles on Research Data Policy and subsequently individual RCUK funding bodies have mandated management of and access to publicly funded research outputs resulting in a flurry of activity in UK higher education institutions to bolster their infrastructure to support RDM. Internationally, publishers are also increasingly mandating that access to underlying research data be made possible to facilitate peer review, validation and reuse.

Case Study: UK funders' requirements for data management and sharing

Since Research Councils UK (RCUK) released its Common Principles on Data Policy (RCUK, 2010) the majority of RCUK funders have mandated the inclusion of data management plans with new grant applications. This approach places the responsibility on individual researchers to provide evidence that data management and sharing issues have

been considered from the earliest stage of research activity. However, RCUK funding body Engineering and Physical Sciences Research Council (EPSRC) took a different approach with the release of their policy framework in March 2011 (EPSRC Policy Framework, 2011). The framework outlines nine expectations relating to research data management and sharing (EPSRC Expectations, 2011) for all UK HEIs seeking EPSRC funding. The expectations are aimed at HEIs rather than the researchers they employ thereby shifting the onus for providing evidence of research data management and sharing infrastructure to the institution rather than the researcher. EPSRC requested that HEIs produce roadmaps outlining how they will meet these expectations by May 1, 2012 and expect institutions to be able to comply fully with these expectations by May 1, 2015. The EPSRC expectations have been a great catalyst for assessing, developing and implementing research data management support systems and have been instrumental for getting the attention of senior management and budget holders within UK HEIs. A number of institutions have made their draft EPSRC roadmaps available via the DCC website (<http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps>).

In addition to expectations pertaining to the management and retention of research data, funders in the UK are increasingly advocating making both research publications and research data openly accessible through gold Open Access (OA) journals. This is clearly evident in the recent changes to the RCUK policy on Open Access (RCUK, 2012) which states that 'Peer reviewed research papers which result from research that is wholly or partially funded by the Research Councils:

1. must be published in journals which are compliant with Research Council policy on Open Access (see section 4).
2. must include details of the funding that supported the research, and a statement on how the underlying research materials – such as data, samples or models – can be accessed.' (RCUK Policy on Open Access to Research Outputs, 2012)

Funding body requirements may seem onerous but it is important to bear in mind that these mandates are aiming to underpin good research practice. Indeed, the notion of data retention as part of good research practice is something that most UK HEIs have already endorsed in response to the RCUK Code of Good Research Conduct which recommended that research data be kept for a period of at least ten years. The DCC provides a handy at-a-glance table summarising a range of UK funders' policies relating to research data management and sharing (<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>). The table is an excellent starting point for anyone wishing to become more familiar with what UK funders expect. The DCC's Sarah Jones provides a more comprehensive overview of a wide range of funders' research data policies, principles and emerging policy trends in chapter three of *Managing Research Data* (Pryor, 2012).

International publishers requirements

Most publishers have offered authors the option of submitting supplementary material such as data along with their publications for some time. However, publishers are increasingly mandating that underlying data be made available both to peer-reviewers and upon request by subsequent readers of the article. Access to the underlying data facilitates validation of the research findings by peers but also enables reuse. The following excerpt from Nature's policy on the availability of data and materials illustrates that data should be retained but also implies that they expect the data to be described sufficiently to allow for validation and ease of reuse.

'An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Therefore, a

condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications (Nature, 2013)

Nature goes further to state that they reserve the right to refuse publication to authors who fail to provide evidence that they are able to comply with the journal's requirements for data sharing. Most recently, FigShare joined up with Public Library of Science (PLOS) in an effort to make the underlying data more readily accessible to 'enable researchers to do more with the data behind the papers, improving the transparency of the research and aid reproducibility' (Digital Science Press Releases, 2013). However, many researchers feel unsure about just what data they need to make available.

Librarians will need to be able to highlight potential benefits associated with research data management activity to secure buy-in from researchers

Unsurprisingly, compliance to funders', institutions' and publishers' policies is not always the best incentive for motivating researchers to spend time – which could otherwise be spent on research – on undertaking data management activities. Heather Piwowar, co-founder of the not-for-profit organisation ImpactStory (<http://www.impactstory.org>), has been carrying out research into the links between making data available and citations and more recently into altmetrics. Her research has revealed that by sharing well managed and curated data, researchers can expect an increase of up to 69% in the number of citations they receive compared with researchers who do not make their underlying data available (Piwowar, 2007). Citations have a direct impact on researchers' career advancement and international standing. With regards to altmetrics, Piwowar states:

'In the next five years, I believe that it will become routine to track — and to value — citations to an online lab notebook, contributions to a software library, bookmarks to data sets from content-sharing sites such as Pinterest and Delicious. In other words, to value a wider range of metrics that suggest a research product has made a difference. For example, my colleagues and I have estimated that the data sets added to the US National Center for Biotechnology Information's Gene Expression Omnibus in 2007 have contributed to more than 1,000 papers. Such attributions continue to accumulate for several years after data sets are first made publicly available.' (Piwowar, 2013)

Over the past few years there has also been a marked increase in the number of data journals available to researchers such as GigaScience (<http://www.gigasciencejournal.com/>) and PANGAEA (<http://www.pangaea.de/about/>). These journals offer an additional means of making research outputs accessible and - more importantly – citable. We have also seen the emergence of a number of negative results journals over the past decade.

'At present, *more than 60% of the experiments fail to produce results or expected discoveries*. This high percentage of "failed" research generates high level knowledge. But generally, all these negative experiments have not been published anywhere as they have been considered useless for our research target'. (The All Results Journal, n.d.)

Negative results journals provide an important opportunity for researchers to share potentially important research findings that may or may not be associated with publications and to get credit for their work. Again, it will be crucial that the data being shared in these journals is well managed and curated to ensure that the negative findings can be picked up and progressed by other researchers.

How do librarians fit into this changing landscape and where do researchers need support most?

Since 2011 the DCC has been working with twenty-one UK HEIs to help them identify the skills and capability gaps within their institutions and to develop and implement research data management policies, support services, and infrastructure (DCC Institutional Engagements, 2012). Based on the outcomes of the initial tranche of engagements, it has become clear that libraries are leading the development and delivery of research data management services in many UK HEIs. Indeed, if the ultimate return on investment for managing, sharing and curating research data is to realise data reuse and ultimately increase the body of scholarly knowledge, libraries have a potentially crucial role to play in facilitating data discovery, contextualisation, assessment and reuse.

Identifying where researchers need support most across the curation lifecycle and the best means of providing that support have been explored through a number of JISC Managing Research Data programme (JISCMRD) and Research Data Management Training Materials (JISC RDMTrain) projects. The DaMaRO project at the University of Oxford recently undertook a survey of researchers in the sciences to find out where they need support most urgently.

‘Perhaps unsurprisingly, the tasks in which most training had been received were those relating to day-to-day management of information: managing bibliographic data; organizing and structuring data within files; storing data securely and backing up; and organizing, structuring, and naming files and folders. Those for which least training was reported were those which concern what happens to data after the end of a project: preparing datasets for long-term preservation; determining whether datasets ought to be preserved; preparing datasets for sharing with other researchers; and dealing with copyright, licensing, or IP issues. It also came as no real surprise to find that the two tasks researchers felt least confident about – dealing with copyright, licensing, and IP issues, and preparing datasets for long-term preservation – were also the areas in which they felt training would be most beneficial’. (DaMaRO Survey, 2012)

What support services should librarians aim to provide?

Librarians have a long history of providing support to researchers in describing and applying metadata to their publications and must now aim to extend their expertise into helping researchers to describe the underlying data. However, description is not the only area where researchers need help. In light of the DaMaRO survey findings, this section will explore some of the additional areas where librarians can help to support researchers in managing and curating their research data in the short and longer-terms. The activities explored in this section are by no mean exhaustive. They represent just few of the many areas where librarians might play an active role as data intensive research becomes increasingly commonplace. The activities listed below include examples of practical activity drawn from ongoing work within the JISC MRD projects, JISC RDMTrain projects, and the DCC’s institutional engagements (IEs).

1. Developing institutional research data policies and support services

As discussed earlier in the chapter, a number of drivers are making research data management and sharing a big priority for UK HEIs. Indeed, the EPSRC requirements on data management have seen a flurry of RDM activity within UK HEIs over the past twelve months with most seeking to develop and implement policies and support services to underpin research data management and sharing. Based on what we’ve seen in the JISC MRD projects and the first tranche of DCC institutional engagements, many UK HEIs are looking to their libraries not just to participate in, but to lead on, the development and implementation of these policies and related support services. However, before an institution

can develop policies or support services that are fit for purpose, reflect working practice and, most crucially, achieve buy in from researchers, an in-depth understanding of current research workflows for different research units and disciplines and an understanding of their data holdings is imperative.

Assessing data holdings and current RDM practice using the Data Asset Framework (DAF)

The Data Asset Framework (DAF) is a freely available survey and interview-based methodology and provides access to a number of sample questionnaires and interview questions. Using DAF can be a great way to get a feel for how much data is being generated within an institution and how it is being managed – or not. DAF has been used by a number of the JISC MRD projects and has been used as a starting point in most of the DCC institutional engagements (IEs). The Jisc-funded Open Exeter project is a good example of RDM policy development and service provision being led by library staff. The Open Exeter project adapted the DAF questionnaires as part of their 'Follow the Data' work and succeeded in getting over 280 respondents from across the institution to feedback on their current research data management and sharing practices. This has provided a sound basis for the project to identify where additional support is needed and how best to implement their research data policy. (<http://as.exeter.ac.uk/library/resources/openaccess/openexeter/data/>).

Start by reviewing what other institutions are doing

A growing number of UK HEIs have made their research data policies (<http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies>) and EPSRC roadmaps available via the DCC website (<http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps>). These can also be an excellent starting point for any institution wishing to get an idea of what might need to be covered in a policy relating to research data as well as tips on how these policies might be implemented and supported. The DCC is also in the final stages of developing a practical how-to guide on setting up and delivering institutional research data management services which will provide case studies based on the Jisc MRD projects and the DCC IEs. The guide will be published in spring 2013.

2. Supporting researchers in their data management planning

As noted earlier, many funding bodies now require researchers to submit a data management plan at the proposal stage. Data management plans (DMPs) should generally be about two to four pages long and cover practical aspects such as the data types, volume, formats, and capture methods that will be used in the project but should also make clear any ethical or intellectual property considerations that may affect data sharing and reuse in the short and longer term. Funders are keen to have evidence that data sharing has been considered and want some information about how and when data will be shared. When dealing with sensitive data or where commercial interests are a concern data sharing may not be possible. In these cases, it will be important to make these restrictions clear in the DMP. DMPs should also make some reference to longer-term preservation and accessibility so it will be important to ensure that researchers' DMPs and impact statements reflect any data protection measures and retention schedules as stated in ethics approval procedures and resulting consent forms. DMPs shouldn't be completed in isolation. They require input from multiple individuals across the institution. Librarians already serve as liaison points between other institutional support services and this experience will be invaluable when helping researchers to collaboratively complete their DMPs. It is important to note here that DMPs should be viewed as living documents. They will – and indeed should - change over

the life of the project and it is important that researchers plan for DMP updates as their projects proceed.

Data management planning using DMP Online

DMP Online was developed by the Digital Curation Centre to enable researchers to build and edit DMPs according to the requirements stipulated by many UK funders. The tool provides researchers with a selection of templates that reflect funder-specific RDM related questions along with tips and guidance for responses. DMP Online encourages users to develop three versions of the plan – a basic version for submission at the bid stage, a detailed version of the plan for the active phase of the project and a final version of the plan which includes publication details and provides information regarding longer term preservation. Several of the institutions that DCC is supporting through their institutional engagements are aiming to develop institutional-specific data management plan templates that reflect local policies and provide access to local guidance and support. In many cases institutions are including named contacts for input on specific questions and developing boiler-plate responses to make the completion of DMPs as straight-forward and efficient as possible. In some cases, for example at the University of Edinburgh and the University of Leeds, the DCC is working with the HEI to integrate institutional versions of DMP Online into existing current research information systems (CRIS) and grant management systems to make data management planning more efficient and to avoid duplicating effort.

3. Providing guidance to researchers for licensing their data

The DaMaRO survey highlighted that licensing and assessing IPR are areas where researchers feel they need greater support. As funders increase their requirements for open access to both publications and underlying research data, the need for this support will continue to grow. Essentially, there are two main options that researchers have when starting out on new research projects. They can apply a license to their data to provide a legally binding statement about what users can – and can't – do with their data or they can decide to apply a waiver for their research outputs which legally states that the author has given up their rights to the works. Creative Commons licenses (<http://creativecommons.org/>) have been increasingly applied to publications and data but many researchers struggle with picking the right license to meet their needs and in many cases go with more restrictive licenses than they need based on uncertainty (van Noorden, 2013). Getting licensing issues pinned down from the outset of the project is essential and it is particularly important to clarify IPR in collaborative research endeavours before selecting any license. As data reuse becomes more commonplace, researchers will also need to be sure that they understand others' data licenses and are able to adhere to any restrictions on sharing and reuse in derived works. The DCC offer a handy how-to guide that explains a wide range of data licenses and provides practical tips on selecting the right one (DCC How-to license research data, 2012).

Make licensing easy for data depositors: licensing data at the Archaeological Data Service (ADS)

The ADS, based at the University of York, requires that data depositors complete a license form (ADS, 2011) that provides a very clear mechanism of making explicit what rights the depositor has with regards to their data, the rights and responsibilities of the repository to undertake curation and preservation activities, and makes clear what end users intending to reuse the data are able to do. The form is concise and helps to ensure that all stakeholders are able to easily identify what they can and can't do with the data. It is vital that HEIs are

equally as clear in any depositor licensing forms that are developed for use within institutional repositories.

4. Helping researchers to select and appraise the data they need to keep

As noted earlier, a key challenge for researchers is identifying what data they need to keep to be able to adhere to publishers', funders' and institutional mandates. The DCC recommends that – as a minimum - researchers keep anything that will be necessary for validating their published research findings. However, in many cases there may be additional data that researchers want to keep and there will also be a number of cases where researchers are legally bound to dispose of the data depending on the conditions outlined in consent forms for research involving sensitive data. It is important to bear in mind that curation requires an investment of resource – whether in the form of effort from researchers during the active phase of the project to value-added services such the application of discovery metadata and preservation actions. The DCC provides a concise overview of seven key criteria that can be very helpful in selecting data worth investment (DCC How to select and appraise research data, 2011).

Make use of the STM and ODE Data publication Pyramid

The International Association of Scientific, Technical & Medical Publishers (STM) and the Opportunities for Data Exchange (ODE) have developed a model to help illustrate the various types of data that may be associated with the findings of a given research publication. All of these, whether born digital, digitised or analogue – need to be considered as part of any research data management plan. The model can be very helpful for librarians when working with researchers to determine exactly what research data outputs they have generated and identifying what must be kept to validate their research findings.

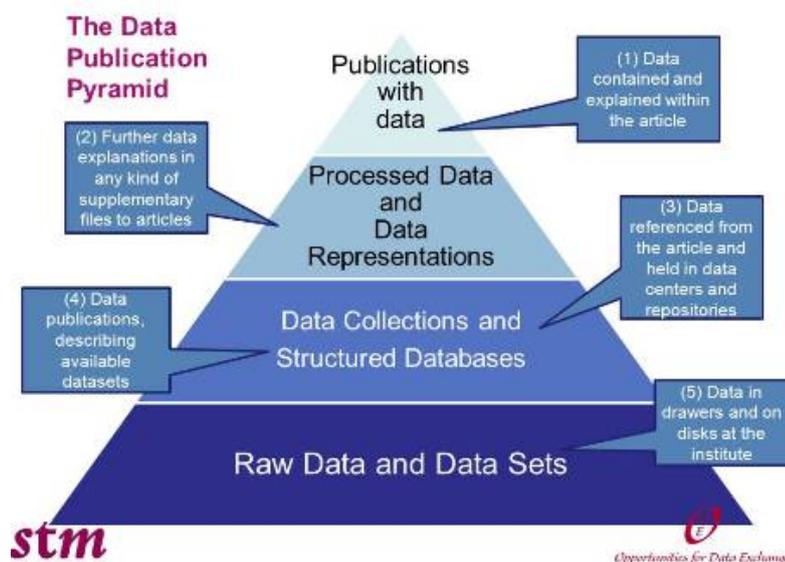


Figure 2. The Data Publication Pyramid (Reilly, 2011)

It is important to remember that selection and appraisal decisions are happening from the earliest stage of new research not just when researchers reach the end of their funded activity. Researchers are constantly making choices about capture tools, formats, compression, software and hardware – all of which might have implications for short and longer term access and reuse.

Active selection: an example from the University of Leeds

Selection and appraisal isn't just something that happens at the point of ingest into a data repository. Selection and appraisal are taking place all the time in research activity. For instance, a researcher at the University of Leeds scans spinal chord images as part of his research. However, the researcher always chooses to capture a larger section of the spinal chord image than he needs to and at a higher resolution than is actually required. This is an active selection decision based on the fact that the most time consuming and costly part of the imaging process lies in preparing the subject for imaging rather than the imaging process itself. By capturing more data than he needs, he is essentially expanding the amount of data available for future research at a fraction of the cost. The additional cost associated with a slightly larger file size is outweighed by the potential savings to be made in carrying out future research and the potential value associated with reuse (RoADMaP project, 2013).

5. Supporting future access and reuse of research data

A major benefit that may be realised from investing in managing and curating research data activity is reuse. As mentioned earlier in the chapter, there has been research showing that researchers who make their data accessible along with publications can expect to see an increase in citations. But, researchers may need some advice on how to make their data citable (i.e., preferred style) and what level of granularity they should use. Researchers may also need support from librarians in citing others' data (i.e., citing external data that has been used as part of new research activity). The DCC has produced a how to cite research data guide that covers many of the main issues that need to be considered (DCC How to cite research data, 2011).

Make data citation simple and clear

Libraries should aim to make citation as easy as possible for researchers and might want to consider automating the generation of recommended citations once data has been deposited into a repository. For example, at the University of Exeter Data Archive (EDA) an automatically generated guide for citations is created once a dataset is ingested into the repository and the researcher receives notification via email. The email contains 'a permanent link to the work and the email urges researchers to 'cite this link in preference to the URL of the item as it provides continuing persistent access in case the URL should ever change' (EDA, 2012). The EDA approach helps to ensure longer-term accessibility to the data as well as making citation easier for reusers.

The short list above illustrates just a few of the possible activities that librarians might become involved in. The LIBER working group on E-Science/Research Data Management's *Ten recommendations for libraries to get started with research data management* (LIBER 2012) provides additional suggestions.

Are librarians ready to provide research data support?

The leadership of the library in establishing and delivering institutional research data management infrastructure and support - as seen in a large number of the DCC's Institutional Engagements - demonstrates that there is recognition of the role and value of librarians' skills. There is also great potential for re-shaping the professional role of the librarian within the institution as these research data management support services mature and become increasingly vital to current and future research. However, to fully realise this vision, there are some areas where librarians may need to adapt and/or update their skills or acquire completely new skills. RLUK commissioned a study in 2012 (Auckland, 2012) which

revealed - via a survey to assess the skills librarians currently feel they need or will need in the near to mid-term - nine key areas where skills gaps might be most significant. The following table lists the nine areas identified in the study and outlines when librarians feel these skills will become essential for their effectiveness in their day to day activities.

The nine areas identified as having potentially the most significant skills gap are

- Ability to advise on **preserving research outputs** (49% essential in 2---5 years; 10% now)
- Knowledge to advise on **data management and curation**, including ingest, discovery, access, dissemination, preservation, and portability (48% essential in 2---5 years; 16% now)
- Knowledge to support researchers in **complying with the various mandates of funders**, including open access requirements (40% essential in 2---5 years; 16% now)
- Knowledge to advise on potential **data manipulation tools** used in the discipline/ subject (34% essential in 2---5 years; 7% now)
- Knowledge to advise on **data mining** (33% essential in 2---5 years; 3% now)
- Knowledge to advocate, and advise on, the use of **metadata** (29% essential in 2---5 years; 10% now)
- Ability to advise on the **preservation of project records** e.g. correspondence (24% essential in 2---5 years; 3% now)
- Knowledge of **sources of research funding** to assist researchers to identify potential funders (21% essential in 2---5 years; 8% now)
- Skills to develop **metadata schema, and advise on discipline/subject standards and practices**, for individual research projects (16% essential in 2--5 years; 2% now) Figure

Figure 3.Reskilling for Research (Auckland, 2012)

Another recent survey was undertaken to investigate current and planned services and support for research data management and bibliometrics across a number of international academic libraries (Corrall et al, 2013). The findings indicate that:

‘...in many cases development of the types of specialized research support services investigated are constrained by knowledge and skills gaps among library staff and a lack of confidence surrounding their expected roles in both RDM and bibliometrics. While technical competencies in both bibliometrics and RDM scored highly among the areas where knowledge and skills were needed, our findings also demonstrate the importance of understanding the research environment at both macro and micro level for providing effective services for research.’ (Corrall et al, 2013)

So, while librarians have many of the core skills required for supporting some areas of research data management, it is clear that additional skills and increased confidence in some areas are considered necessary by many professionals in the field.

How can librarians re-skill to provide research data management support?

We've seen an increase in the number of educational programmes covering aspects of data management and curation being offered to emerging information science professionals in recent years (<http://www.dcc.ac.uk/training/data-management-courses-and-training>). Sheila Corrall provides a very good summary of emerging educational programmes covering RDM for information science students in *Managing Research Data* (Pryor, 2012). There have also

been some recent developments with regards to professional development opportunities for librarians already in the workplace. The following examples may be of particular interest for professional development.

DC101, Tools of the Trade, and Train the Trainer materials

The DCC offers free, half day courses that provide 'an introduction to research data management and curation, the range of activities and roles that should be considered when planning and implementing new projects, and an overview of tools that can assist with curation activities' (DC101, 2012). These introductory courses aim to give participants an understanding of the drivers and benefits for research data management and help support staff to identify practical ways to develop and sustain RDM support within their current institutional infrastructure. The DCC also offers half day workshops on a range of DCC tools including the Data Asset Framework (DAF) and DMP Online. These courses provide hands-on experience in applying the tools in an institutional setting. The DCC also provides access to a range of DC101 and JISC RDMTrain materials to help support staff develop and deliver RDM training within their own institutions.

3TU Data Intelligence 4 Librarians

The 3TU Federation is a consortium of the libraries of the Delft University of Technology, Eindhoven University of Technology, and the University of Twente. The consortium provides a shared data service for researchers across the three institutions. In 2011, 3TU develop their Data Intelligence 4 Librarians course to help progress the 'professionalisation and positioning of the library and its employees to be a partner in the support of data-intensive science' (3TU, 2011). The course provides a mix of remote and face-to-face training and exercises designed to help librarians in post develop and hone RDM skills. Topics include data citation, selection, and legal issues. The course is run over four day long face-to-face session and students undertake assignments in pairs as sharing experience and knowledge is a core tenet of the course. Trainers and fellow students provide feedback on the course work. In 2012, the course was launched online.

RDMRose

RDMRose is 'a JISC funded project to produce taught and continuing professional development (CPD) learning materials in Research Data Management (RDM) tailored for Information professionals' (RDMRose, 2012). The project is being run collaboratively by the libraries of the Universities of Leeds, Sheffield, and York and will make use of the Information School at the University of Sheffield to disseminate the eight RDM modules – each of which equates to a half day of study. 'One of the central assumptions made in the design of the module is that librarians themselves often do not have in-depth experience of research. RDM and an increasing number of other roles to support research require more understanding of the perspective of the researcher. Therefore considerable time in the module is devoted to actively exploring the nature of research and research data. The module also encourages you to think about the potential role of other professional services, such as research administration and computing' (RDMRose, 2012). The modules are also being embedded into the

postgraduate education for information professionals so the skills of emerging and in post librarians will be of a consistent quality.

The Digital Curator Vocational Education Europe (DigCurV) project has been working to develop a vocational training framework for library, archives, and museum staff. The project has also developed and maintains a registry of current international training opportunities that may be of value to librarians seeking to undertake professional development (<http://www.digcur-education.org/eng/Training-opportunities>). The Library of Congress' Digital Preservation Outreach and Education (DPOE) also provides access to a list of current training opportunities for different audiences from beginners to experts (DPOE Calendar, 2013) and runs a number of train the trainer workshops (DPOE Train the Trainer, 2012) to improve librarians' skills.

Conclusions

Francis Maude, Minister for the Cabinet Office and Paymaster General refers to data as 'the 21st century's new raw material' (Open Data White Paper, 2012). Indeed, those who can claim to be data scientists are currently 'difficult and expensive to hire and, given the very competitive market for their services, difficult to retain' (Davenport, 2012). However, as dealing with data of all kinds - not just research data but administrative and financial data – is becoming increasingly ubiquitous in all walks of life, employers across all fields will be seeking recruits who have the ability to exploit and extract value from data. As noted by Hal Varian, Google:

'The ability to take data - to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids.' (Varian, 2009)

As Davenport says, 'think of big data as an epic wave gathering now, starting to crest. If you want to catch it, you need people who can surf' (Davenport, 2012). This changing landscape does pose a number of challenges for librarians as they carve out a niche in a world of data. However, librarians have risen to these sorts of challenges before and are well placed to capitalise on the new opportunities that are opening up. So, grab your surfboard cause surf's up!

Guide to abbreviations

DCC	Digital Curation Centre
DMP	data management plan
HEI	higher education institution
IE	institutional engagement
OA	open access
RDM	research data management

References:

Auckland, M. for RLUK, (2012) <http://www.rluk.ac.uk/files/RLUK%20Re-skilling.pdf>

Archaeology Data Service and AHDS Archaeology Licence Form (2011)
<http://www.ahds.ac.uk/documents/ahds-archaeology-licence-form.doc>

Corrall, S., Kennan, M., and Afzal W. (to be published May 2013), Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research, *Library Trends*, 61(3).

Birte Christensen-Dalsgaard et al, Ten recommendations for libraries to get started with research data management, Final report of the LIBER working group on E-Science / Research Data Management (2012)
<http://www.libereurope.eu/sites/default/files/The%20research%20data%20group%202012%20v7%20final.pdf>

DaMaRO Research Data Management Training for the Sciences Survey (2012),
<http://blogs.oucs.ox.ac.uk/damaro/2012/11/21/damaro-survey-results-research-data-management-training-for-the-sciences/>

Data Intelligence 4 Librarians (2012) <http://dataintelligence.3tu.nl/en/home/>

Davenport, T.H. and Patil, D.J. (2012) Data Scientist: The Sexiest Job of the 21st Century, Harvard Business Review, <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1>

DC101 (2012) <http://www.dcc.ac.uk/training/digital-curation-101>

DCC (2013) Curation Lifecycle Model <http://www.dcc.ac.uk/digital-curation/what-digital-curation>

DCC (2012) DMP Online <https://dmponline.dcc.ac.uk/>

DCC (2012) How-to guides <http://www.dcc.ac.uk/resources/how-guides>

DCC (2012) Institutional Engagements <http://www.dcc.ac.uk/community/institutional-engagements>

DCC (2013) Overview of funders' data policies at <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

DCC (2012) Tools of the Trade workshops <http://www.dcc.ac.uk/training/tools-of-the-trade>

DCC (2012) Train the Trainer <http://www.dcc.ac.uk/training/train-the-trainer>

DCC (2013) What is Digital Curation <http://www.dcc.ac.uk/digital-curation/what-digital-curation>

Digital Science Press Releases (2013) <http://www.digital-science.com/inside-ds/press-releases#figsharePLOS>

EPSRC Expectations (2011)
<http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>

EPSRC Policy Framework on Research Data (2011)
<http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx>

Exeter University Repositories
<http://as.exeter.ac.uk/library/resources/openaccess/repositories>

Guy et al, Pinning It Down: Towards a Practical Definition of 'Research Data' for Creative Arts Institutions (2013) as presented at IDCC13

Higgins, S. (2008) The DCC Curation Lifecycle Model, *International Journal of Digital Curation*, 3 (1), 134-40.

Higgins, S. The lifecycle of data management in Managing Research Data (2012) Pryor G, editor. Facet Publishing, ISBN 978-1-85604-756-2

European Research Area (ERA) http://ec.europa.eu/research/era/optimal-circulation_en.htm

Jisc Managing Research Data Programme (JISCMRD), (2013)
<http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>

Jisc news release (2013), <http://www.jisc.ac.uk/news/stories/2013/01/jisc-apc.aspx>

Jisc Research Data Management Training Materials (JISC RDMTrain), (2013)
<http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain.aspx>

Jones, S. Research data policies: principles, requirements and trends in Managing Research Data (2012) Pryor G, editor. Facet Publishing, ISBN 978-1-85604-756-2

Library of Congress' Digital Preservation Outreach and Education (2012)
<http://www.digitalpreservation.gov/education/index.html>

MaDAM into Sustainable Service (MiSS) project (2013)

Nature Policy on Availability of data & materials
<http://www.nature.com/authors/policies/availability.html>

OECD Principles and Guidelines for Access to Research Data from Public Funding, OECD 2007, <http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf>

Open Data White Paper: Unleashing the Potential (2012)
http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf

Open Exeter project (2012) <http://as.exeter.ac.uk/library/resources/openaccess/openexeter/>

Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3): e308. doi:10.1371/journal.pone.0000308

Piwowar, H. Altmetrics: Value all research products (2013) *Nature* 493, 159
doi:10.1038/493159a

RCUK Policy on Open Access to Research Outputs (2012)
http://www.rcuk.ac.uk/documents/documents/RCUK%20_Policy_on_Access_to_Research_Outputs.pdf

RCUK Common Principles on Data Policy (2010)
<http://www.rcuk.ac.uk/research/Pages/Datapolicy.aspx>

RDMRose project (2013) <http://rdmrose.group.shef.ac.uk/>

Reilly, S et al. Report on Integration of Data and Publications (2011) [http://www.stm-
assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf](http://www.stm-
assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf)

Research Data Canada (2013) [http://rds-sdr.cisti-icist.nrc-
cnrc.gc.ca/eng/about/backgrounder.html](http://rds-sdr.cisti-icist.nrc-
cnrc.gc.ca/eng/about/backgrounder.html)

Researchers of Tomorrow: the research behaviour of Generation Y doctoral students (2012)
[http://www.jisc.ac.uk/media/documents/publications/reports/2012/Researchers-of-
Tomorrow.pdf](http://www.jisc.ac.uk/media/documents/publications/reports/2012/Researchers-of-
Tomorrow.pdf)

RoaDMaP project, University of Leeds (2013) <http://library.leeds.ac.uk/roadmap-project>

The All Results Journal (2013) <http://www.arjournals.com/ojs/>

University of Glasgow Draft Research Data Policy (2012)
http://www.gla.ac.uk/media/media_232425_en.docx

University of Glasgow Code of Good Practice in Research, Version 1.2, (2011)
http://www.gla.ac.uk/media/media_227599_en.pdf

University of the Arts Research Data Management Policy (2012)
[http://www.arts.ac.uk/media/research/documents/UAL%20Research%20Data%20Managem
ent%20policy.pdf](http://www.arts.ac.uk/media/research/documents/UAL%20Research%20Data%20Managem
ent%20policy.pdf)

Van Noorden, R. Researchers opt to limit uses of open-access publications (2013) Nature
[http://www.nature.com/news/researchers-opt-to-limit-uses-of-open-access-publications-
1.12384](http://www.nature.com/news/researchers-opt-to-limit-uses-of-open-access-publications-
1.12384); doi:10.1038/nature.2013.12384

Varian, H. on how the Web challenges managers, The McKinsey Quarterly January 2009
[http://www.mckinseyquarterly.com/Hal_Varian_on_how_the_Web_challenges_managers_22
86](http://www.mckinseyquarterly.com/Hal_Varian_on_how_the_Web_challenges_managers_22
86)