Getting Data Creators on Board with the Digital Curation Agenda

Lessons Learned in Developing Training for Researchers

Dr Meriel Patrick and Dr James A. J. Wilson
DaMaRO Project
University of Oxford
Oxford, UK
meriel.patrick@it.ox.ac.uk

***Abstract*—University research projects are a key source of digital information with potential long-term value. Researchers rarely need to be persuaded that preserving the fruits of their work is in principle a good thing, but may often lack knowledge of the best way to go about doing this. Additionally, time pressures on academics are such that curation can frequently end up being pushed down the priority list. It is therefore important that information professionals working alongside researchers are able to offer appropriate training and advice on both the practicalities of and the rationale for digital curation.**

The DaMaRO Project is one of a series of research data management projects based at the University of Oxford. The project's remit includes developing training for researchers (intended to encourage them to consider data sharing and preservation issues at an early stage in their research), plus the development of an institutional data archive (DataBank) and catalogue of datasets (DataFinder). This paper offers some reflections on our experiences thus far, and in particular looks at the question of how researchers and others who are involved in the creation of digital data may most effectively be engaged in planning for and facilitating its long-term preservation.

Keywords*—*Research data, research data management, digital curation, data creators, researchers, training, universities, HEIs.

I. Introduction

University research projects are an important source of digital information with potential long-term value. Academic researchers can collect and generate vast quantities of data in the course of their work, and as data are frequently susceptible to a wide range of different types of analysis, interpretation, and comparison, it is rare for any single research project to fully exploit the potential of a given dataset.

However, while researchers will usually wholeheartedly agree that the data they produce are a valuable resource, they are not always fully aware of the most effective and appropriate means of preserving those data, and practical barriers can often stand in the way of the data being made available for future use.

This paper reports some of the findings of work done at the University of Oxford over the last few years, and offers some reflections on the sort of training and advice that could usefully be offered by information professionals working alongside researchers, with the aim of encouraging and facilitating the effective curation of research outputs.

The work at Oxford has concentrated chiefly on research data management; therefore, the focus of this paper is on the preservation of digital research data, rather than other aspects of digital curation.

II. Research Data Management Work at the University of Oxford

Over the past four years, a series of projects focusing on research data management have been undertaken at the University of Oxford. The work has been cross-departmental, involving input from IT Services, the Bodleian Libraries, Research Services, and the academic divisions. Three key projects in the series are Sudamih (Supporting Data Management Infrastructure for the Humanities, 2009-11), VIDaaS (Virtual Infrastructure with Database as a Service, 2011-12), and most recently DaMaRO (Data Management Roll-out at Oxford, 2011-13)[50].

All three projects have included a training strand, intended to encourage researchers to take a closer look at their data management practices throughout the research lifecycle. An important aspect of this is consideration of what happens to data at the end of a research project: how they can best be preserved, and made available for others to use. Training activities to date have included:

- Two half-day courses for humanities researchers
- A half-day course currently being offered to all four of Oxford's academic divisions
- Training events run in collaboration with the Digital Curation Centre
- Various shorter training events and presentations, offered through individual divisions or departments
- Development of resources for use in researcher induction sessions

[50] Sudamih and DaMaRO are JISC-funded projects. VIDaaS was funded by HEFCE and JISC as part of the University Modernisation Fund.

- Contributions to the University of Oxford's central Research Data Management website
- Contributions to the University of Oxford's Research Skills Toolkit website
- Leaflets and fact-sheets for researchers

Other project activities have included contributing to the development of the University of Oxford's Policy on the Management of Research Data and Records (formally adopted in July 2012), and development of software tools which will ultimately form part of Oxford's research data management infrastructure. Three major tools that are emerging as part of this process are ORDS (the Online Research Database Service), designed both to aid researchers in working with active research data and to facilitate easy archiving at the end of the project; DataBank, which will be the University of Oxford's institutional data archive; and DataFinder, which will provide a catalogue of Oxford datasets held in DataBank, ORDS, and elsewhere.

To inform the work being undertaken, the projects have also engaged in requirements gathering, exploring researchers' knowledge of and attitudes to a range of research data management issues, and seeking their views on the type of services and training they would like to see provided. This was done through a mixture of face-to-face interviews and online surveys. The findings of the Sudamih and VIDaaS Projects are detailed in their respective Researcher Requirements Reports, both available online. [1] [2].

Information gathering during the DaMaRO Project has included two surveys, both of which took place in late 2012. The first focused specifically on research data management training for researchers working in the sciences[51]; the second was open to all University of Oxford researchers, and looked at research data management practice and awareness more generally[52]. The results of these surveys are available from the DaMaRO website [3] [4].

III. Key Challenges and Possible Solutions

A. *Researchers' Attitudes to Data Preservation and Sharing*

The surveys and interviews conducted in Oxford have tended to focus more on data sharing than on curation or preservation considered in the abstract. However, preservation and sharing are often closely associated in researchers' minds: there is a widespread assumption that if data are being deposited in an archive or repository, or otherwise prepared for long-term storage, this is chiefly for the purpose of making them available for re-use, either immediately or after an embargo period. Hence it is often difficult to separate researchers' attitudes to data curation from attitudes to data sharing.

The Oxford work indicates that many researchers who create or collect data are not averse in principle to sharing these. However, in practice, a number of factors may prevent this from occurring, or at least make it more problematic.

A 2011 survey of researchers, run as part of the VIDaaS Project, revealed that a large majority (85%) of respondents felt that a substantial portion of their data were of potential value or interest to other researchers in higher education. Almost two thirds (63%) said that this also extended to people outside the HE community. However, only 41% reported that they would be happy to make their own research data available once they had completed the work they intended to do and published the results. Even fewer than this (34%) had previously published data.

The 2012 Oxford RDM survey painted a slightly more encouraging picture. Thirty percent said they would be prepared to share all or most of their data (possibly after an embargo period), and another 40% were willing to share at least some of them.

Practical bars to data sharing include the following:

- Lack of awareness of appropriate places to deposit data
- Lack of knowledge of appropriate way to present material for long term preservation
- Concerns about the risks of sharing data too early
- Lack of time to prepare or deposit data
- Ethical and legal issues
- Financial issues

B. *Lack of Awareness of Appropriate Places to Deposit Data*

The Oxford RDM survey asked researchers whether they had ever deposited data in a dedicated repository or data store. Those who had not (61% of respondents) were asked why this was. The most popular reason (given by almost exactly half of the respondents who had not deposited data) was simply that they did not know of an appropriate place to put it.

This lack of awareness is not limited to data repositories. The same survey also asked researchers whether they had heard of or used a number of tools and services, which included ORA, the University of Oxford's institutional repository for textual research outputs. Almost half (47%) had never heard of it, and most of the rest (a further 41%) had never actually used it.

This is an area in which additional training could clearly be of value: improving researchers' knowledge of repositories and archives is a straightforward way to remove a significant bar to preservation of research data.

However, in some cases, training may need to do more than simply direct researchers to the appropriate archive. A number of respondents in the Oxford RDM survey commented that the sheer size of their datasets (which may be on the terabyte scale) made sharing difficult. In these cases, researchers may also need guidance on data selection, and on appropriate technologies for data storage and transfer. There may also be a need for further development of infrastructure capable of dealing with these volumes of data.

[51] Hereafter referred to as the DaMaRO science training survey.

[52] Hereafter the Oxford RDM survey.

C. *Lack of Knowledge of Appropriate Way to Present Material for Long Term Preservation*

Even when researchers know where to deposit data, they may not always know how to prepare it. Some respondents to the Oxford RDM survey commented that they did not know how to put their data into an appropriate format: this seemed to mean more than just not knowing which file formats to use, but extended to uncertainty over the best way to present data to make them re-usable.

This issue also surfaced in the DaMaRO science training survey. This asked researchers about eleven key data management tasks: questions covered their level of confidence, the quantity of training that they had received, and how useful they felt additional training would be. A clear picture emerged: the respondents generally had lower confidence about and expressed a greater desire for training in those aspects of data management which related to long-term preservation.

The four specific areas in which respondents felt additional training would be most useful were:

- Dealing with copyright, licensing, or other IP (intellectual property) issues relating to datasets
- Preparing datasets for long-term preservation
- Data documentation
- Preparing datasets for sharing with researchers outside their research group

D. *Concerns About the Risks of Sharing Data Too Early*

Researchers are often reluctant to share data before they have been comprehensively mined for publications. In most fields, considerable emphasis is still placed on traditional research outputs such as journal articles and monographs: publication of datasets is not yet regarded as having the sort of value – either in terms of contribution made to the community of knowledge, or in terms of the benefit to an individual researcher's academic reputation – that traditional publications have. Releasing data into the public sphere at too early a stage is therefore seen as a risky enterprise, as it raises the possibility that another researcher may draw similar conclusions and publish first.

A respondent to the Oxford RDM survey expressed this common concern:

> "There is risk of getting 'scooped', so given the current funding climate (which is heavily based on publication record), until research is published, I would be hesitant to freely share data."

Even when publication of initial results has already occurred, if there are further conclusions to be drawn from the data, many researchers would prefer to have the opportunity to do this themselves. This sometimes leads to a perception of researchers as data hoarders, selfishly hugging material to themselves rather than allowing others to make use of it – something which is viewed with particular disapproval if the research that produced the data was publicly funded. However, researchers frequently find themselves in a difficult position: they are under pressure from funding bodies to demonstrate that they have made good use of the funding, and from their institutions to produce research outputs which will count towards the REF[53]. Although data publication may count for something (and indeed is increasingly being required by funding bodies – see, for example, the Research Councils UK Common Principles on Data Policy [5]), it remains the case that researchers' worth is measured chiefly by means of traditional research publications.

A comment from a senior history researcher, interviewed during the Sudamih Project, reflects the tension that many academics feel:

> "In principle, you want material to be available, and I believe in sharing. On the other hand, if you've just spent five or ten years collecting a dataset and you haven't yet milked it for what it's worth, and you've had funding to do the project, then you're very nervous about handing over that dataset."

Another Oxford RDM survey respondent made a similar point:

> "[I have] Often not finished getting the most out of my data, want to be able to return to it at a later date and not have someone else publish my data (which has happened)."

E. *Lack of Time to Prepare or Deposit Data*

A related point concerns the time and effort required to prepare datasets for preservation and sharing. Data gathering is itself often a lengthy process, and hence researchers tend only to collect what is necessary for the specific purpose they have in mind. Data that have been collected for personal use are thus often untidy and incomplete. They may also employ idiosyncratic standards, and make use of abbreviations or conventions that would require considerable explanation to be intelligible to other users. In some cases, researchers may store the raw data and their own private notes and comments together (in the same database, for example), and the latter may need to be removed before the former can be published.

The process of making datasets fit for public consumption can therefore frequently be an arduous one. A number of researchers interviewed during the Sudamih Project commented that they would need to do significant further work on their data to get them into a state in which they would be happy to publish them. This would inevitably take time away from other academic endeavours, and given the priority of written outputs noted above, there are currently few incentives for researchers to do this.

A comment from an Oxford RDM survey respondent expressed a similar view:

> "Part of my reluctance to share data is that my data is fairly roughly organised, and in various stages of polishedness (recordings, transcriptions, etc.), so it would

[53] The Research Excellence Framework: the exercise by which the research of British higher education institutions will be assessed, scheduled to take place in 2014.

be quite a big project to get it all presentable, and I'm not sure in what format I would do it."

Additionally, as the deposit of data tends to happen as a project draws to a close, it is also easy for it to get overlooked or pushed to one side in the rush to complete everything on time.

There is some scope for training to address this problem: if researchers can be encouraged to have data preservation in mind from the beginning of a project, there is a greater chance that the data will be collected, organized, and documented in a way that will make them more accessible to subsequent users, thus reducing the need for a large amount of reworking at the end of the project, and increasing the likelihood of a final dataset that the researcher is willing to share.

However, if this point and the one above are to be fully addressed, it is important to take researchers' concerns about data sharing seriously. Training needs to cover not simply the practicalities of preserving research outputs, or even ways of making this more straightforward (although these are both important topics), but also the rationale for doing so – that is, it needs to consider the *why* as well as the *how*. If this is to be effective, it needs to focus not just on researchers' obligations (to share data in order to meet funding bodies' requirements, for example), but also on the benefits to the individual researcher (such as increased potential for citations), and on the usefulness of data to the research community at large.

While the widespread tendency to view datasets solely as a means to an end rather than as valuable research outputs in their own right persists, it is likely that curation of those datasets will continue to be viewed as a lower priority. To rectify this, a major cultural shift in attitudes is needed. While pressure from funding bodies and institutions may help to spur this process on, such a change will ultimately come only from the researchers themselves – from a paradigm shift in perceptions about the value of alternative types of research outputs.

There are some pockets of the research community in which data is already recognized as a valuable resource in its own right. High-energy physics, for example, is an area involving a large number of very specialized roles, and where the contribution of scores of people may be necessary to gather the data required to support a single written research output such as a journal article. It is not uncommon for high-energy physics papers to credit two or even three hundred authors; many of these individuals will not have been actively involved in the writing of the paper, but will instead have worked only on the generation of the dataset underpinning it. Their contribution is nevertheless recognized, valued, and consequently credited. Areas of research that operate in this way are also more likely to have well established processes for preserving and sharing data, and these have typically been initiated from within the research community as a necessary tool to facilitate effective work, rather than imposed from outside as a result of funding bodies' requirements.

Although other areas of research function very differently, the example of high-energy physics provides reason to hope that, given time and appropriate encouragement, a similar culture can be fostered elsewhere. Researchers who have gained a greater appreciation of the benefits of data curation are more likely to engage in it, resulting a greater quantity of high-quality datasets being available for re-use by the academic community. This may even ultimately result in a positive feedback loop, where the perceived value of data preservation (and thus the motivation to ensure it happens) increases as the benefits of having more data available become clear.

Training is, of course, only one part of the picture. There is also a pressing need for further work to be done on lowering the barriers to curation, by providing intuitive, easy-to-use tools and processes that are straightforward to integrate with researchers' existing workflows. Training can and should, however, form an important part of an interim solution, by drawing researchers' attention to the services that already exist, and by advising on ways to make the process as smooth as possible.

F. Ethical and Legal Issues

It is not uncommon for researchers to find themselves unable to share some or all of their data as a result of ethical concerns (generally relating to confidentiality, and appropriate consent from research subjects), or legal issues relating to data ownership, especially when datasets have been supplied by third parties.

While it is clearly desirable for confidential information to be suitably protected, there is a strong case to be made for encouraging researchers to give careful consideration to the permissions they ask for when obtaining consent from research subjects. In some cases, data with significant potential re-use value may have to be kept private (or even destroyed), not because the subjects of the research were unwilling for data to be shared, but simply because they were never asked. Researchers working with sensitive data may thus benefit from guidance on how they can meet their responsibilities to their subjects without unnecessarily restricting data use. This may include advice on appropriate wording for consent forms, and on anonymization of datasets intended for wider dissemination. (The UK Data Archive provides a range of helpful resources for researchers on this and related topics: see, for example, [6].)

G. Financial Issues

Finally, financial issues are a consideration for some researchers. Long-term preservation of data comes at a cost, and it is not infrequently the researchers themselves who have the responsibility of securing funding for this. The University of Oxford's DataBank service, for example, is likely to be run on a cost recovery basis for projects with more than a very small amount of data to preserve. While it is hoped that it will ultimately be possible to secure central University funding to cover the cost of storing data from unfunded research, in the short term, it is probable that the service will only be able to accommodate data that come with funding attached.

This means that the long-term curation of data may often need to be factored into project budgets and grant applications – which in turn means that it needs to be planned for from the very earliest stages of a project, before the research itself has even begun (at least one respondent in the Oxford RDM survey stated that data had not been deposited because this had not been budgeted for in the grant, and potential solutions proved too expensive). Researchers need to be made aware of the various options that are available to them, and of the costs attached to these. Funding bodies also vary in the extent to which they are prepared to pay for long-term curation, so researchers may also need guidance on this front.

IV. Practical Aspects of Training

A. *Nature and Format of Training*

Researchers are busy people with many calls on their time, and while they may acknowledge in theory that digital curation is an important topic it would be helpful to know more about, in practice, attending training about it often comes a long way down the list of priorities. It is therefore important that face-to-face training sessions are kept relatively brief and well focused, and that written guidance material is concise and easy to navigate.

Opinions were divided regarding the relative usefulness of face-to-face training courses and online or print materials: both were acknowledged to have advantages and disadvantages. Paradoxically, a chief advantage of online training materials – the fact that they are available to be consulted at any time – was also viewed as a disadvantage, in some cases by the same researchers who had cited this as a benefit. It was noted that the fact that online training can be done at any time often leads to it not being done at all; face-to-face training, on the other hand, requires researchers to set aside a specific time period for the course, and having done this, they are then likely to spend that time focusing on the topic under consideration.

It also seems that researchers use face-to-face training and online guidance in different ways: the former is more likely to be sought out by those who want an overview of the subject as a whole, whereas written guidance is often used when researchers are seeking an immediate answer to a specific question or problem that has arisen in the course of their work. There is thus a strong case for having both available where possible.

B. *Timing of Training*

During the course of the work in Oxford, it has become very clear that digital curation cannot be viewed simply as something to be bolted on to the end of the research process. If consideration is not given to how data will be preserved from an early stage in a project, it is substantially less likely that this will happen at all: data may be in an inappropriate format, or lacking documentation, or there may be a lack of budgetary provision for long-term storage, or researchers may simply run out of time before data can be prepared and deposited.

Training provision therefore needs to reflect this: guidance needs to be available to researchers at all points in the research lifecycle. The Oxford interviews revealed a general consensus that it would be useful for researchers to receive initial training relatively early during their time as graduate students. However, it was felt that the best time was not right at the beginning of the course, but after a few weeks or months – perhaps during the second term. This was for two reasons: first, because students often find themselves overwhelmed with information when beginning a new course, and secondly, because once graduates have spent a little time engaged in research, they have a better idea of the issues they are likely to face, and thus have a clearer idea of how to apply what they learn to their own work.

C. *Content of Training*

With regard to the content of training, comments from attendees at Oxford courses have indicated that researchers find it extremely helpful to have concrete examples to illustrate the points under discussion. Digital curation involves concepts that may be unfamiliar to researchers (“ingest”, “metadata”, and “migration”, for example), and these can often be best conveyed by demonstrating what they might look like in a real-world situation.

Fig. 1 and Fig. 2 below are word clouds generated from participant feedback from two research data management training events held in Oxford. Attendees were asked how the courses could be improved. The size of the words is proportional to their frequency of occurrence, and in both cases clearly shows the demand for more examples.

D. *Choice of Language in Training Materials*

A question from the Oxford RDM survey also highlighted the importance of using language that is familiar to researchers. The question asked whether respondents had ever deposited research data “in a dedicated repository or data store”, and was intended to elicit whether researchers had made arrangements for the long-term preservation of their research data. Many researchers understood it this way, and answered accordingly, reporting that they had deposited data with the UK Data Archive, the Dryad Digital Repository, the Archaeology Data Service, and a range of similar bodies. However, at least a third of the respondents interpreted the question as asking about their day-to-day arrangements for storing active research data: answers included departmental or research group shared storage, the University of Oxford’s central back-up service, Dropbox, and even external hard drives. This emphasizes the need for clarity, and an awareness that key terms such as “repository” may not conjure up the same set of associations for all parties.

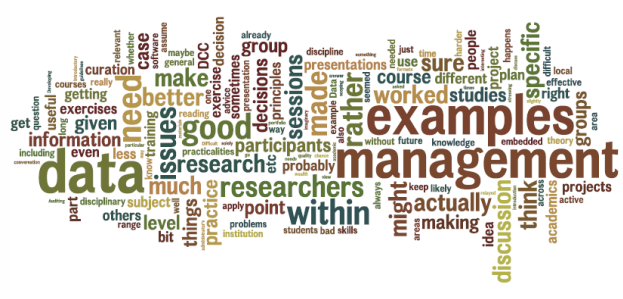

Figure 1. Word cloud generated from participant feedback from a 2010 Digital Curation Centre workshop in Oxford

Figure 2. Word cloud generated from participant feedback from a 2013 DaMaRO Project training course

Even the use of the word "data" can be controversial in some circumstances. While scientists and social scientists are generally happy to use this term to describe the information that supports the conclusions of their research, researchers in the humanities (particularly those working with less structured sources) may not think of themselves as data users – though they may nevertheless have gathered collections of material with significant potential for re-use. This can make it more difficult to reach these researchers, as they are likely to regard training or guidance material which mentions research data as irrelevant to their work.

Finding a suitable alternative term has proved challenging, however. The Sudamih Project attempted to address this problem by creating materials for humanities researchers which referred instead to "research information". While this was well received by the researchers, it unfortunately caused confusion in another direction, as this phrase is now widely used to refer to administrative information *about* research (details of the nature and scope of projects, for example, or of the number and type of research outputs).

The current Oxford approach is to use the term "research data", but to immediately qualify this by defining "data" as broadly as possible. Where appropriate, other more humanities-friendly phrases such as "source materials" and "information used in research" may also be included. This is not a complete solution, however, as researchers who do not think of themselves as data users may not read beyond the title of a course description or guidance document, and hence may never see the broader definition. The debate about terminology is ongoing.

E. Formal Vs. Informal Training

The DaMaRO science training survey questions distinguished formal training – for example, taught courses, online modules, or exercises on which feedback was given – from informal, which might take the form of on-the-job training or advice from colleagues or supervisors, or being directed to electronic or print guidance or information. The distinction was made partly in response to some anecdotal evidence suggesting that in the sciences, training for new researchers is more likely to be delivered informally, in labs and research groups. The survey responses confirmed this: for all eleven tasks covered, respondents had received significantly more informal training than formal.

This has implications for the way in which training is delivered. One option is simply to attempt to increase the quantity of formal training available. However, if a culture already exists in which key research skills are fostered informally, it may be beneficial to try to tap into this. This might involve targeting training resources at supervisors, rather than (or in addition to) new researchers themselves, and ensuring that adequate guidance material is freely available and well publicized to both senior researchers and more junior ones.

Engaging supervisors and senior researchers in training provision has multiple possible benefits: it provides a potential route for reaching large numbers of researchers, and over time, would aid the process of embedding the principles of digital curation in academic culture. However, the DaMaRO Project's experience suggests it is likely to be difficult to achieve on a large scale, not least because many senior researchers currently lack the relevant expertise. It is also hard to measure the extent and effectiveness of informal training. This being the case, it seems wise to attempt a multi-pronged approach, and to provide material intended to encourage informal training alongside more traditional formal training resources.

F. Overall Tone of Training

It has already been noted that a key driver of digital curation is funding bodies' requirements that research data be made publicly available at the end of research projects. It is clearly desirable that researchers should be properly informed about their obligations, and there is little doubt that having requirements imposed externally is an effective means of getting researchers to comply.

However, it is important that training does not focus on this to the exclusion of all else as the motivation for digital curation. When many researchers already feel over-worked and under pressure, piling on yet another obligation can all too easily result in an attitude of weary resentfulness. Researchers are not rebellious schoolchildren who need to be bullied into

working harder: they are generally highly motivated and highly skilled individuals who take a great deal of pride in what they do, and thus are more likely to embrace digital curation as a worthy goal if persuaded of its merits.

A fine example of this approach is provided by a leaflet produced jointly between the DICE, SHARD, and PrePARe Projects[54]: this outlines the benefits of research data preservation (plus the skills necessary to achieve it), under the title 'Sending your research material into the future'. A PDF of the leaflet is available online [7].

The half-day training workshop currently being offered to Oxford researchers as part of the DaMaRO Project begins by summarizing the University's Policy on the Management of Research Data and Records. As this focuses chiefly on researchers' responsibilities, this may at first sight appear to be an obligation-based approach rather than a benefit-driven one. However, an institutional policy is a document that serves multiple purposes: in addition to setting out what is expected of members of that institution, it also provides a statement of the institution's values. Drawing researchers' attention to the University's policy – which states in its opening sentences that research data are valuable – sends a strong message that their institution regards data as an important resource, and that caring for and preserving them appropriately constitutes a crucial part of good academic practice.

There is, however, something to be said for placing more of an emphasis on the requirements of external bodies such as funders in discussions *about* training – that is, in proposals for new training courses, or when arguing the case for the inclusion of digital curation training as part of an existing curriculum. In such circumstances, highlighting the financial considerations can often provide a swift and effective way of stressing the importance of the topic, and may thus be helpful in securing institutional buy-in.

V. Summary and Conclusion

If research data are to be effectively curated, the creators or compilers of those data need to be engaged with the curation process from the earliest stages of a project. The work at the University of Oxford indicates that researchers often need guidance regarding the practicalities of curation – where and how to deposit data, and the best format in which to present them. In some cases, they may also need some encouragement to regard data curation as a worthwhile activity – something that is of sufficient value to merit taking time away from other academic endeavours.

Training and guidance needs to be available to researchers throughout the research process, starting from an early stage in their careers. As researchers have many calls on their time, training should be kept relatively concise, and ideally offered in multiple formats (e.g. face-to-face courses plus online materials) to provide a measure of flexibility. Researchers have a definite preference for material with a practical focus, using familiar language, and offering specific concrete examples. Finally, it is important that training does not dwell too much on researchers' obligations and the penalties for failing to meet them, but that it also emphasizes the benefits of digital curation: its chief aim should be not to threaten, but to inspire.

[54] All three of these projects were funded by JISC.

References

[1] J. A. J. Wilson and M. Patrick, "Sudamih researcher requirements report," 2010, http://sudamih.oucs.ox.ac.uk/docs/Sudamih Researcher Requirements Report.pdf.

[2] M. Patrick, "VIDaaS researcher requirements report", 2011, http://vidaas.oucs.ox.ac.uk/docs/VIDaaS%20Researcher%20Requirements%20Report.pdf.

[3] M. Patrick, J. A. J. Wilson, and P. Jeffreys, "DaMaRO Project survey on research data management training for scientists – results", 2012, http://damaro.oucs.ox.ac.uk/docs/RDM%20for%20sciences%20-%20training%20survey%20results.xlsx.

[4] J. A. J. Wilson, P. Jeffreys, M. Patrick, S. Rumsey, and N. Jefferies, "Results of the 2012 University of Oxford research data management survey", 2013, http://damaro.oucs.ox.ac.uk/docs/OxfordRDMsurvey 2012_public.xlsx.

[5] Research Councils UK, "Common principles on data policy", http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx.

[6] Van den Eynden, L. Corti, M. Woollard, L. Bishop, and L. Horton, Managing and Sharing Data, 3rd ed., UK Data Archive, 2011, pp. 22–27, http://data-archive.ac.uk/media/2894/managingsharing.pdf.

[7] DICE, SHARD, and PrePARe Projects, "Sending your research material into the future", 2012, http://lsedice.files.wordpress.com/2012/07/dice-shard-prepare-leaflet.pdf