Research Data Management Training for Support Staff

A DaMaRO Project Survey

Meriel Patrick (DaMaRO Project), Dorothy Byatt (DataPool Project),
Federico De Luca (DataPool Project), Wendy White (DataPool Project),
Sally Rumsey (DaMaRO Project), and James A. J. Wilson (DaMaRO Project)

Table of contents

Introduction

In March and April 2013, the DaMaRO Project ran a survey for University of Oxford staff whose role involves supporting researchers. The survey was intended to gauge support staff's awareness of and confidence regarding a range of research data management topics, and to ascertain the areas in which training might be helpful.

The survey was run in collaboration with the DataPool Project at the University of Southampton. DataPool had previously run a similar survey for Southampton support staff, and the Oxford survey was based on this (with a few alterations to reflect the Oxford context). The survey was also hosted on Southampton's iSurvey system, and members of the DataPool Project provided considerable assistance with the analysis of the survey results.

Survey respondents

The Oxford survey attracted thirty-seven valid responses, mostly working in IT provision, the libraries, or research support. The exact breakdown is given below:

Respondents' department or role	Number	Percentage of total
IT Services	8	21.6%
Libraries – Subject Librarian	5	13.5%
Libraries – Other	4	10.8%
Research Services	7	18.9%
Divisional or Departmental Research Support	12	32.4%
Doctoral Training Centre	1	2.7%
Total	37	100%

Respondents were also asked if they had personally been involved in carrying out research, either as a researcher or as part of an advanced postgraduate research degree. Slightly under two-thirds (62%) had. These were divided roughly equally between those who had been researchers, those who had done a research degree, and those who had done both.

Responses and observations

The survey asked support staff about a number of different tasks relating to research data management. These were grouped into categories:

Planning for data management
- Use of file naming to assist file management and retrieval
- Version control of files so that it forms good practice
- Identifying Intellectual Property (IP), copyright, data protection, and other legal issues relating to research data
- Identifying ethical issues relating to research data, for example, issues of confidentiality
- Requirements of specific funders' to make research data available for re-use
- Writing data management plans for submitting with bids/research proposals to funders
- Compliance with the University of Oxford's Policy on the Management of Research Data and Records.

Data management during the research process
- Identifying the data storage requirements – size, type – during the lifetime of the project
- Awareness of data security issues during the lifetime of the project
- How to describe the data so that others on the project can find the data, i.e. adding basic working metadata
- How to describe the data so that others in the future are able to find and safely reuse the data, i.e. adding public metadata

Post-research data management
- Options for the dissemination and sharing of data
- Issues associated with the licensing of the data for re-use
- Issues associated with the longer-term preservation and archiving of data
- Identifying the costs associated with the creation of data during the project including storage, dissemination and preservation
- Requirements of specific funders to allow the exploitation and re-use of data

For each task, respondents were asked:

- How confident they personally felt about their knowledge and ability to handle a query on this topic
- How confident they felt about being able to refer researchers to the right person, organization, department, or resources for advice on the topic (in this section, respondents were also given the option of indicating that they considered advising about this task to be their role)

For each task, respondents were asked to rate their confidence level on a scale from one (not confident) to seven (completely confident).

Mean confidence levels were as follows:

	Confidence in advising	Confidence in referring elsewhere[1]
Planning for data management		
Use of file naming	3.76	3.89
Version control	3.62	4.09
IP, copyright, data protection, and other legal issues	3.51	4.76
Ethical issues	3.84	4.47
Funders' requirements regarding making data available for re-use	3.84	4.39
Writing data management plans	2.65	3.85
Compliance with the University of Oxford's Policy	2.89	3.79
Data management during the research process		
Identifying data storage requirements	2.70	3.44
Data security issues	3.05	3.65
Describing data for others on project – basic working metadata	3.05	3.71
Describing data for future re-use – public metadata	2.89	3.61
Post-research data management		
Dissemination and sharing of data	3.35	3.94
Licensing of the data for re-use	2.81	3.81
Longer-term preservation and archiving of data	2.78	3.74
Identifying costs associated with the creation of data	2.22	3.37
Funders' requirements to allow the exploitation and re-use of data	3.22	3.70

Overall levels of confidence thus seem to be low to moderate. Respondents did generally seem to be somewhat more confident regarding their ability to refer researchers elsewhere for advice, which is perhaps an encouraging sign; however, no task had a mean confidence level of greater than 4.76.
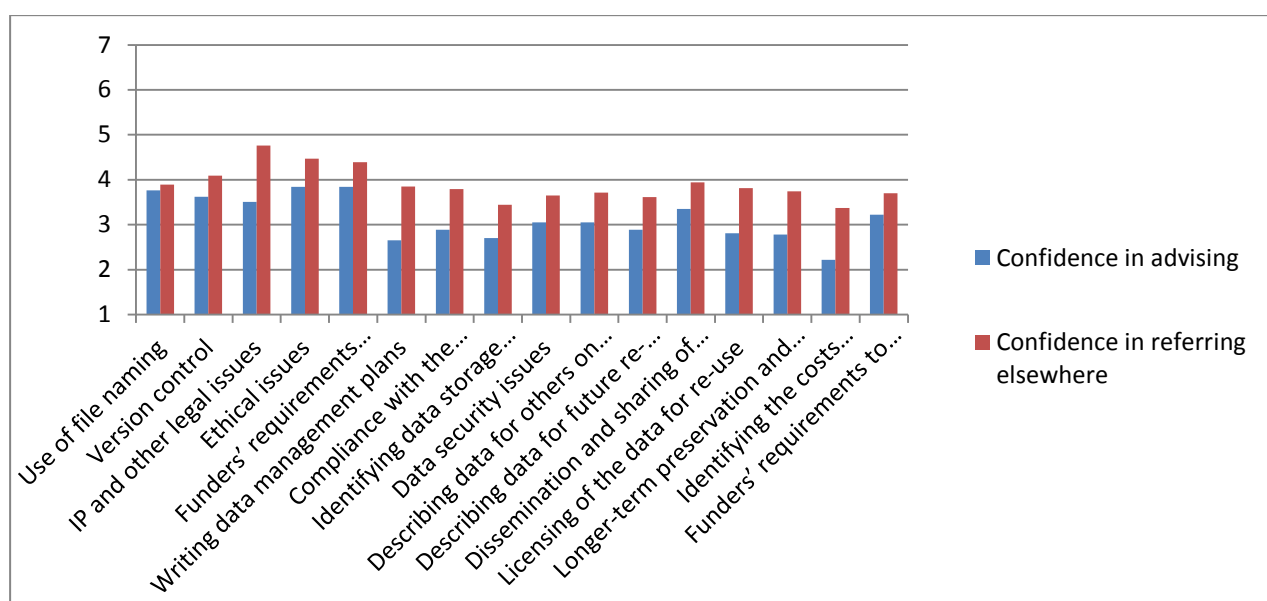

Figure 1: Bar chart comparing mean confidence values

[1] Confidence level for referring queries elsewhere excludes respondents who specified that they considered advising on this task to be their role.

Tasks sorted by mean confidence value (least confident to most):

Mean confidence in advising researchers	Mean confidence in referring researchers elsewhere for advice
Identifying costs associated with the creation of data	Identifying costs associated with the creation of data
Writing data management plans	Identifying data storage requirements
Identifying data storage requirements	Describing data for future re-use – public metadata
Longer-term preservation and archiving of data	Data security issues
Licensing of the data for re-use	Funders' requirements to allow the exploitation and re-use of data
Compliance with the University of Oxford's Policy	Describing data for others on project – basic working metadata
Describing data for future re-use – public metadata	Longer-term preservation and archiving of data
Data security issues	Compliance with the University of Oxford's Policy
Describing data for others on project – basic working metadata	Licensing of the data for re-use
Funders' requirements to allow the exploitation and re-use of data	Writing data management plans
Dissemination and sharing of data	Use of file naming
IP and other legal issues	Dissemination and sharing of data
Version control	Version control
Use of file naming	Funders' requirements regarding making data available for re-use
Ethical issues	Ethical issues
Funders' requirements regarding making data available for re-use	IP and other legal issues

As noted above, respondents were also given the opportunity to indicate whether they considered advising on a given task to be part of their current role. Responses were as follows:

Task	Number	Percentage
Planning for data management		
Use of file naming	2	5.4%
Version control	3	8.1%
IP, copyright, data protection, and other legal issues	4	10.8%
Ethical issues	5	13.5%
Funders' requirements regarding making data available for re-use	8	21.6%
Writing data management plans	3	8.1%
Compliance with the University of Oxford's Policy	2	5.4%
Data management during the research process		
Identifying data storage requirements	1	2.7%
Data security issues	1	2.7%
Describing data for others on project – basic working metadata	1	2.7%
Describing data for future re-use – public metadata	1	2.7%
Post-research data management		
Dissemination and sharing of data	2	5.4%

Licensing of the data for re-use	5	13.5%
Longer-term preservation and archiving of data	2	5.4%
Identifying costs associated with the creation of data	2	5.4%
Funders' requirements to allow the exploitation and re-use of data	10	27.0%

A separate question asked whether respondents felt they *should* know who to refer researchers to for advice on each task, regardless of their current confidence in being able to do so.

Task	Number	Percentage
Planning for data management		
Use of file naming	17	45.1%
Version control	21	56.8%
IP, copyright, data protection, and other legal issues	30	81.1%
Ethical issues	29	78.4%
Funders' requirements regarding making data available for re-use	30	81.1%
Writing data management plans	26	70.3%
Compliance with the University of Oxford's Policy	31	83.8%
Data management during the research process		
Identifying data storage requirements	16	43.2%
Data security issues	18	48.6%
Describing data for others on project – basic working metadata	15	40.5%
Describing data for future re-use – public metadata	16	43.2%
Post-research data management		
Dissemination and sharing of data	28	75.7%
Licensing of the data for re-use	24	64.9%
Longer-term preservation and archiving of data	23	62.2%
Identifying costs associated with the creation of data	24	64.9%
Funders' requirements to allow the exploitation and re-use of data	27	73.0%

Thus the proportion of respondents who considered that advising on any given task formed part of their role was generally fairly small. By contrast, for most of the tasks listed, a majority of respondents considered that they should know who to refer queries to. A notable exception to this trend was tasks that relate to data management during the research process: for these, only a minority (though a fairly substantial one) felt that they needed to know where to refer researchers for advice. These tasks also had the smallest percentage of respondents who considered that advising on them formed part of their role, suggesting that the survey respondents generally felt that planning and post-research data management fell more squarely within their remit than active data management during the research process.
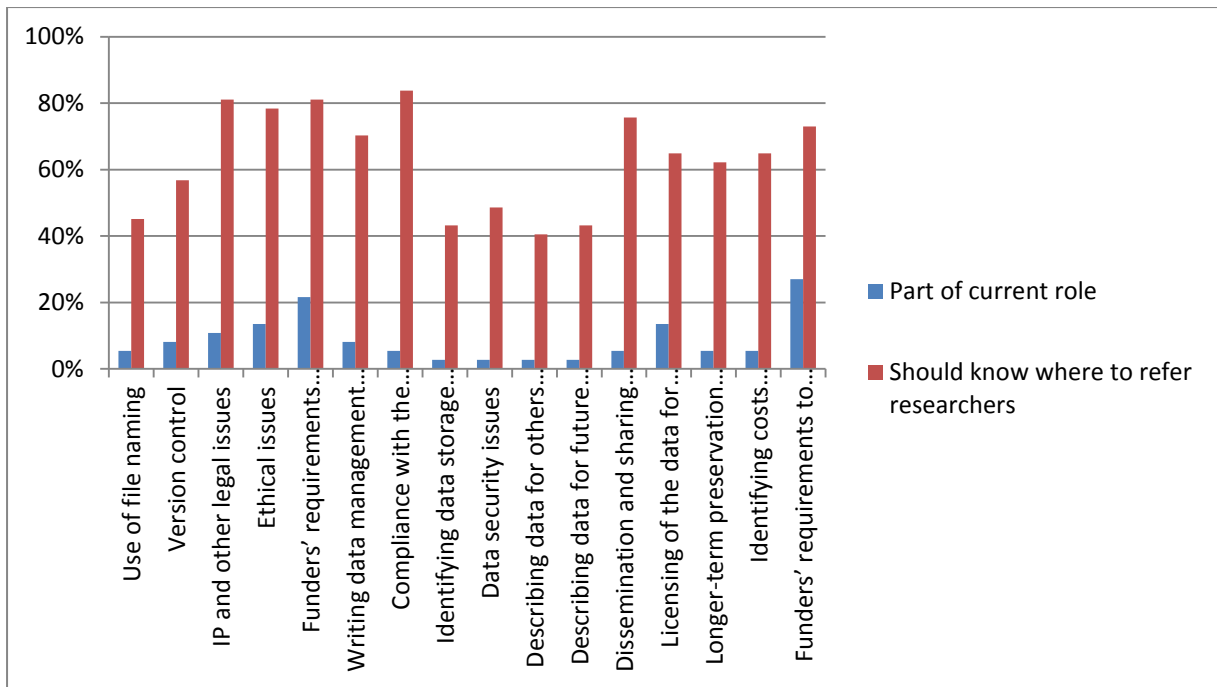
Figure 2: Bar chart comparing respondents' perception of their role and the need to know where to refer researchers

In the final section of the survey, respondents were asked whether they had been asked for help related to research data management in the last six months. Responses were as follows:

Frequency	Number	Percentage
Not at all	17	45.9%
Once only	6	16.2%
Two to five times	11	29.7%
More than five times	3	8.1%
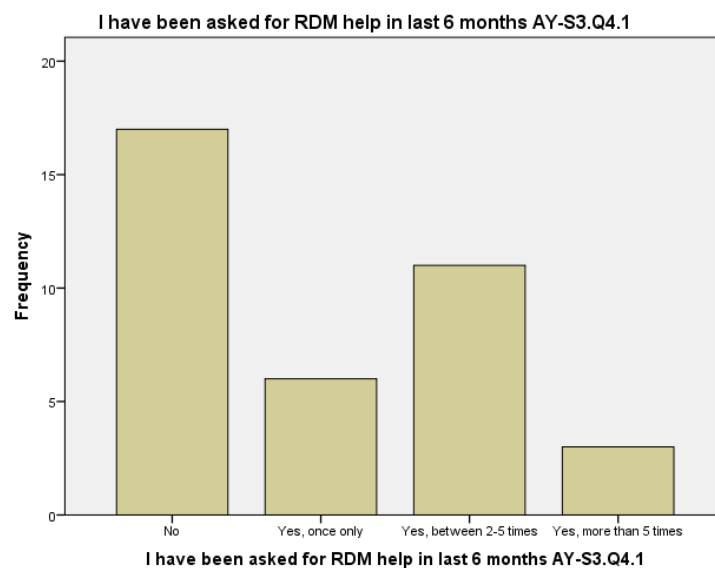
Figure 3: Bar chart showing frequency of requests for RDM help (chart produced as part of analysis by DataPool staff)
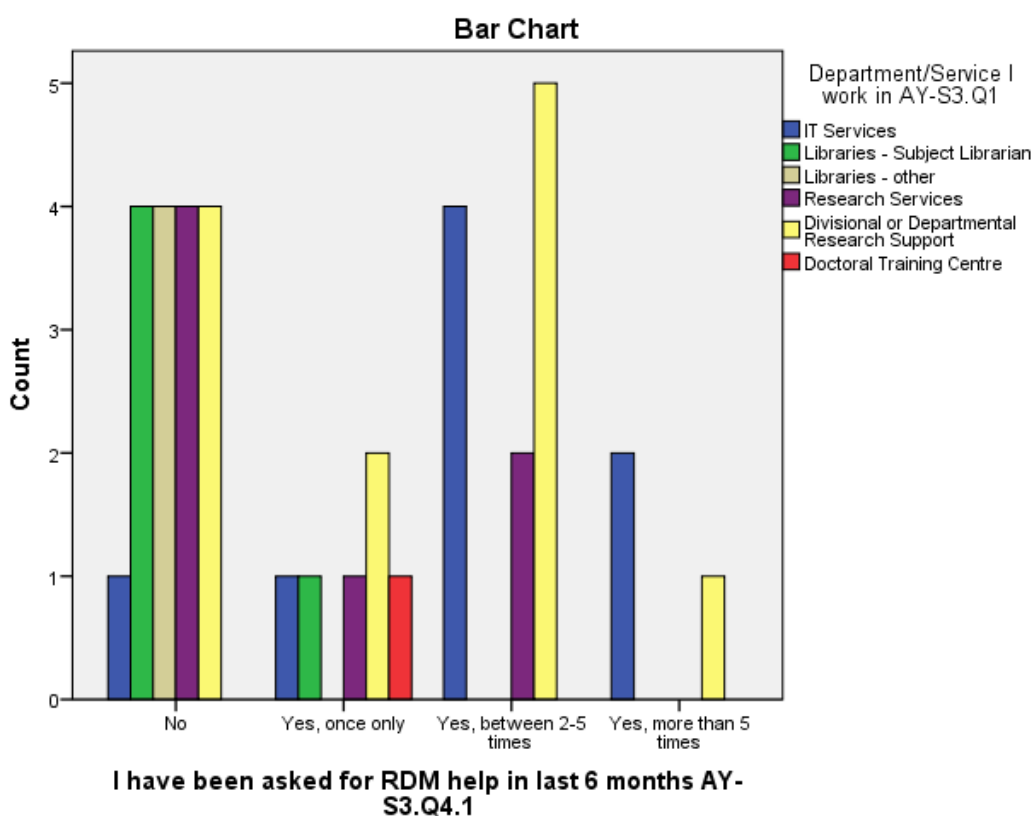
Figure 4: Bar chart showing frequency of requests for RDM help, broken down by respondents' role (chart produced as part of analysis by DataPool staff)

Figure 4 shows responses broken down by respondents' role or department. It is noticeable that respondents from the libraries received fewest requests, whereas respondents who received a larger number of requests were more likely to be working in IT Services, divisional or departmental research support or (to a lesser extent) Research Services.

Respondents were also asked about their priorities for training – specifically, which area they considered training most essential for. Responses were as follows:

Task	Number	Percentage
File management - naming and versioning	2	5.4%
Data Management Plans	8	21.6%
Legal issues and research data	7	18.9%
Ethical issues and data	0	0.0%
Making data available for re-use	4	10.8%
Security and storage of data	2	5.4%
Describing your data	2	5.4%
Funder requirements for research data	6	16.2%
Data management costs	1	2.7%
Determining whether datasets should be preserved	0	0.0%
Where and how to archive research data	4	10.8%
Other[2]	1	2.7%

---

[2] This option invited respondents to specify an area for training. However, the one respondent who selected this option chose not to do so.
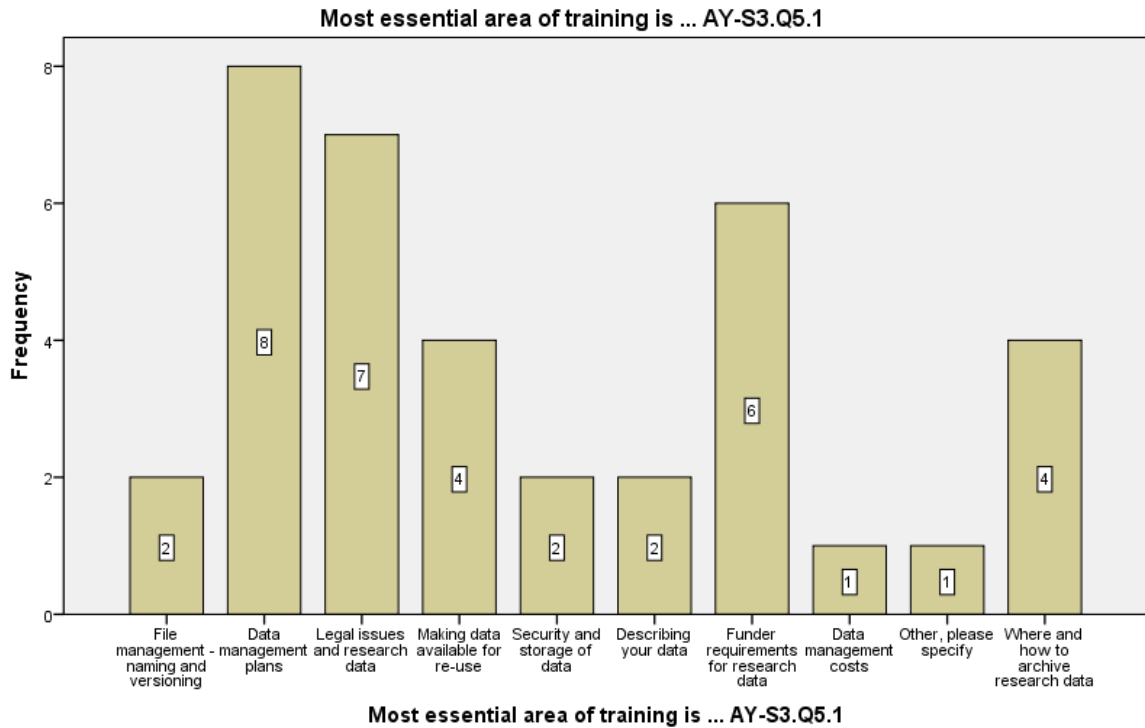
Figure 5: Bar chart showing area in which training was deemed most essential – options with zero answers excluded (chart produced as part of analysis by DataPool staff)
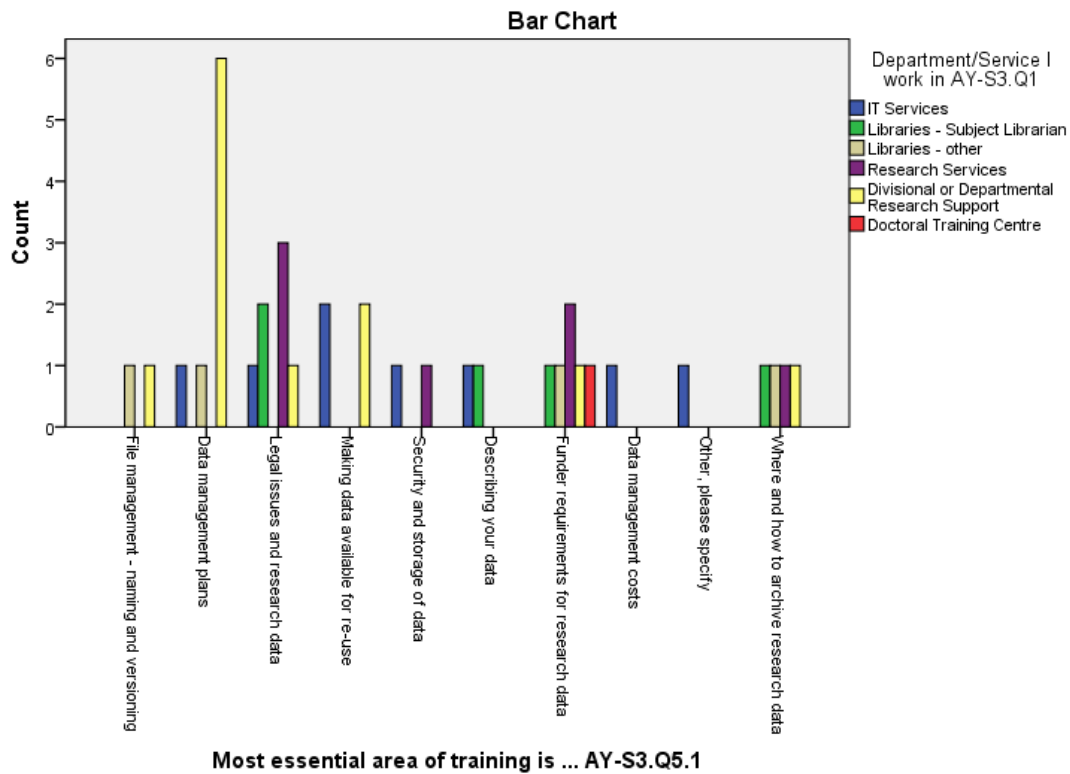

Figure 6: Bar chart showing area in which training was deemed most essential, broken down by respondents' role – options with zero answers excluded (chart produced as part of analysis by DataPool staff)

Figure 6 shows responses broken down by respondents' role or department. It is chief notable for the comparatively large demand for training in producing data management plans from respondents working in divisional and departmental research support.

Free text comments

In each section of the survey, at least one free text field was provided to allow respondents to add further comments, or to highlight areas where training would be particularly welcome.

Many comments reflected a general need for more training or guidance in this area:

- "I'd need guidance on every single topic."
- "I'm not very aware of who I should be directing these questions to."
- "Vital to know who to direct researchers too for all of these areas!"
- "I know it's important for a researcher to have a data management plan, but I don't really know where to start advising on how to make one."
- "Training on all areas would be extremely useful please."
- "I would greatly appreciate as much training and support as is available."

Some respondents commented more specifically on the role of library staff on advising on research data management:

- "Research data management is an area which is poorly understood outside specialist and expert groups of staff. Front-line library staff are intimidated by much of the unfamiliar terminology. Wider explanation of the basic concepts would be very useful to broaden the knowledge base of staff across the University Libraries."
- "I think research data management is an area where subject librarians could make significant contributions but that at the moment the training and resources do not exist to allow us to begin working in this area."

A few respondents also offered more substantive comments on what was needed:

- "I hope there's a way to ease academics into this... it feels quite heavy and fussy if I may say that. I think it is important to not appear dictatorial and to make sure we provide a seamless system/solution rather than just lots of advice/you must do xyz."
- "The deficits are not in training but in staff resources and services to support researchers. We've got plenty of expertise around the University but hardly any staff effort devoted to it."
- "I think this is a matter for the data protection office and IT Services to manage. Ideally they would produce boiler-plate data management plans, generic costs for storing data and clear guidance for PIs. In other words I should know where to turn to for this specific information rather than knowing it myself."

Conclusions

The relatively small sample size limits the extent to which firm conclusions can be drawn from this survey. However, some tentative observations may be made.

In general, levels of confidence in both advising and referring researchers regarding data management tasks are lower than seems desirable: ideally, all support staff would have a high level of confidence in advising on those aspects of research data management most closely related to their role, and would be at least reasonably confident in directing researchers to the right place for advice on other topics. This suggests a need to make additional training and advice available to staff who are involved in supporting researchers. Particular areas of focus might include identifying the costs involved in data management and identifying data storage requirements (both topics respondents were particularly lacking in confidence about), plus data management planning, legal issues, and funder requirements (all identified by respondents as priorities for training).

It is interesting to note that for any given task, only a small proportion of respondents regarded advising on it to form part of their role. While it not necessary for all support staff to be able to advise on all aspects of data management, this does suggest a need to ensure that adequate advice and assistance is available to researchers – and just as importantly, that researchers know where to find it. Similarly, the fact that only a minority of support staff feel it falls within their remit to know where to refer researchers for further advice on tasks relating to management of data during the research process highlights a need to ensure that there are no major gaps in the support provided to researchers.

Finally, the spread of requests for assistance with research data management is also of interest. It is hard to know how representative these results are, but among this group at least, requests for help with RDM were far more likely to be directed to IT Services or research support staff than to library staff – despite the fact that a significant proportion of current research data management work (particularly that relating to post-research long-term preservation) is currently based in the libraries. This may argue for a need to raise the profile of library staff as a source of information about research data management, or to make doubly sure that adequate advice is available from the other sources – or, ideally, both.