WP2 Requirements Analysis (initial draft questions for semi-structured interviews), Preliminary report, July 2012

The RoaDMaP requirements analysis work package (WP2) aims to build on work done in other projects to inform our own approach to working with our RoaDMaP case studies. Project staff drew together questions from a number of sources to inform our approach to semi-structured interview – listed in the References section at the end of this document. It is clear from talking to other JISC projects and data management colleagues that some found a structured approach to interviewing helpful, whilst others abandoned this approach in favour of a much more informal, unstructured approach.

The interview questions aim to :

(i) aid the project's understanding of how researchers create, store, use and access data, their attitudes towards sharing data and their current and anticipated support needs,

(ii) provide a framework for face to face interviews with case researchers linked to the RoaDMaP case studies,

(iii) provide a framework for fact to face interviews with researchers outside the RoaDMaP case studies, either by RoaDMaP staff or additional staff funded by the institution

Discussions at our project meetings in March and May 2012 suggested the questions were too detailed. The main sections of the interview were agreed: these are largely based on the DCC's Data Management Planning headings. Some more 'broad brush' questions were suggested and discussed (highlighted in green on the template). It was felt the broader questions did not reflect all aspects of the research data lifecycle – for example, the questions implied data generation was a one off activity rather than reflecting the analysis and manipulation of data performed by those generating it.

We agreed the way forward was to trial the questions with interviewees and hone them as we went along, structuring the interviews around the broad agreed headings and choosing follow up questions depending on interviewee responses and their particular areas of interest.

The semi-structured interview comprises (i) a project information sheet and consent form sent in advance (ii) a face to face interview of about 1 hour, recorded digitally (if consent given). It is anticipated that more detailed follow-up discussions on specific areas of data management may be required; the project is considering how these follow up discussions will be approached. Questions for use with research support staff who may not be directly involved in generating data are also being considered.

Comparison exercise

In the table below the basic interview questions in the green shaded boxes have been compared against the detailed set of questions.

Requirements Analysis Detailed & Basic Questions for researcher data interview

	1 Detailed questions	2 Basic questions	3 Optional follow on questions
1	**Researcher and research context**		
a.	**What is your name?**	1. Personal details and role.	
b.	**What is your job title?**		
c.	**What is your research role? Are you part of a research group?**		
d.	**Tell me about the research you do.**	2. Tell me about the research you do.	
e.	**Who funds your research?**	3. Does your research funder have requirements relating to data (DMP, locus of deposit, longevity of access)? 4. The University RDM policy requires a data management plan for each proposed research project or funding application – do you already do this? What support would help?	**How long are you required to preserve your data for beyond the life of the research? (We'll come back to this).** **Do you have any collaborators on your research i.e. UK based, European or international? Re collaborators do you know what DMP arrangements they have in place?**
f.	**Does your funder have any data management requirements? What are they?**		
g.	**Are you aware of data management planning and if so have you ever produced a dmp**		
2	**Data Types, Formats, Standards and Capture Methods**		
a.	**What is the exact nature (range, scope) of your research data?**	5. What sort of data do you produce and how is it stored in the short term. Any issues?	**How would you describe your data i.e. observational, experimental, documents, computer code, reference, derived, simulated, other.** **Can you elaborate on the different data stages, short, medium and long term processes that exist once data is acquired i.e. at the Raw Data, Organised data, Analysed data stages.** **What tools – software or hardware – are used in generating the data? And in utilizing the data?**
b.	**In what format(s), will you store your data in the short term after acquisition?**		
c.	**Do you know the approximate size of the data / number of files?**		

d.	**Once data is stored is it static or dynamic?**	6. Is the data static or dynamic? 7. Looking to the future, do you see the research data you generate changing in any way: increasing/decreasing in volume, changes in format etc.?	
e.	**How is your data organised, described and annotated?**	8. Do you create documentation and description of the data? Is it enough to allow reuse (by you, by someone else)?	**What level of data description do you capture i.e. title and short textual description and what medium will it be captured in?** **Does metadata documentation exist that describes your research data sets?** **How will these descriptive metadata be created or captured?**
f.	**Is this amount of organization and description sufficient for another person with similar expertise to be able to understand and properly use the data? (Come back to metadata later in interview)**		**Are there any issues with the way you organise your data; what could be improved? What benefits would this give?**
3	**Ethics and intellectual property**		
	Who owns the data arising from your research, and the intellectual property rights relating to them?	9. Who owns the data you generate?	**Is your data bound by any privacy or confidentiality concerns?** **Are you aware of any contractual arrangements that exist within your research, agreements or otherwise?** **What is your understanding of copyright and licensing issues and how have they been incorporated?**
4	**Access, Data Sharing and ReUse**		
a.	**With whom will you share your research data in the short term?**	10. Do you share your data with others? Would you? Why / why not?	
b.	**Beyond your immediate colleagues, what data do you would be willing to share and with whom? How soon would you share it?**		**If not able/willing to share, why not?** **Are there any access issues re your data that would cause you concern** **Do any enhanced security controls need to exist? Would the disclosure of the project data breach the Data Protection Act?** **What would motivate you to share your data?** **Have you ever re used your own data from previous projects?**

c.	**If participants are involved in your research, what consent arrangements exist?**		
d.	**In the journals, or places you publish most often, are data, or other supplemental information, accepted for publication?**		**How should the link between data and publications be maintained?**
5	**Short term storage and data management**		
a.	**How's your data stored and backed up in the short term?**		
b.	**Have you ever experienced storage problems due to the size of the files?**		
c.	**Who is responsible for the immediate day-to-day management, storage and backup of the data arising from your research?**		**Who will be responsible for your research once your research has concluded?**
d.	**Do you discuss your practical data requirements (short term and longer term) with Faculty IT Staff? (When?)**		
6	**Deposit and long-term storage (questions asked depend on depth of discussion in Access, Data Sharing and Re-use)**		
a.	**Do you know your storage requirements currently, short, medium, long term?**		
b.	**Have you ever deposited your data in a repository and, if so, which and why?**	11. What happens to the data longer term? Do you deposit to a data centre or similar?	**At what stage did you submit your data? Who deposited it? (Or, if never deposited, who would you envisage depositing the data)?** **Any comments about the deposit process or accessibility of the data? Do you receive any usage data from the repository / would you like to?** **For how long will you embargo your research data before it is published for others to see and use?** **If your research data is held in a repository, is the data then deleted from its original storage area** **How do you keep track of where your data is stored and what it relates to?** **Have you ever lost any data and/or do you currently have any data which is no longer compatible with existing software or on hardware media that are not now widely readable?**

c.	**Who might be interested in your data and what value would it have for them?**		**Does the data potentially have a current or future commercial value to either the University or funding body?** **How do you imagine that people would find your data set?**
d.	**Under what data-sharing license will you publish your research data?**		
e.	**What persistent identifiers will be used to permit correct citation of your datasets?**		
f.	**What metadata will be published with the data to make them interpretable and reusable?**	See 8.	**Are you aware of any metadata standards widely adopted in your discipline?**
g.	**Who will decide which of your research data are worth preserving?**		**Do any arrangements exist to meet the funders preservation requirements** **How long would your data set be useful or have value for you or others if it were to be preserved?**
7	**Resourcing**		
a.	**Have you included the relevant costs of data management requirements in your research bids?**		
b.	**Who will be responsible for your data, once you have left your present research group?**		
8	**Training and Support**		
a.	**For all members of staff with access to project data, do the staff members need training regarding the criticality and sensitivity of the data that is being handled and can you evidence that suitable training has been or will be undertaken?**		
b.	**Have you ever received any training around research data management and planning?**		**Was the research data training helpful?** **What other types of support would you or your colleagues like to see?** **Who should provide research data management training?**
c.	**Can you think of any examples of good data management practice in your own work or that of others?**		**Is this good practice relevant for all researchers? Can we incorporate it into our training?**

9	Your discipline		
a.	**Within your research discipline are there key points of note regarding research data and it's processing that you would like to raise?**	12. What else should we be asking you about your research data?	
b.	**Thinking about the topics discussed, are there any key changes you'd like to see in your own practice or in how the University support you? What benefits would these changes offer?**		
		13. Who else would be useful to interview about research data management?	

References

Shotton, D (2012) Twenty Questions for Research Data Management, http://datamanagementplanning.wordpress.com/2012/03/07/twenty-questions-for-research-data-management/

University of Leeds / Russell, M (2011) Research Support Process Functional Specification , ISS.

Purdue University Libraries, Data Curation Profiles Toolkit - Interview worksheet. http://datacurationprofiles.org [accessed 24/04/2012]

Digital Curation Centre (2011) Checklist for a Data Management Plan http://www.dcc.ac.uk/webfm_send/431

Whitton, Michael and Takeda, Kenji (2011) Data management questionnaire results: IDMB Project. Southampton, GB, University of Southampton, 16pp. http://eprints.soton.ac.uk/196243/

University of Exeter (2012) Open Exeter Research Data Survey http://blogs.exeter.ac.uk/openexeterrdm/files/2012/04/survey-questions1.pdf

Research360, (2012) Managing data across the institutional research lifecycle: Data Management Plan for PGRs v0.2, University of Bath. http://blogs.bath.ac.uk/research360/2012/03/postgraduate-dmp-template-first-draft/

Archaeology Data Service, (2011) DataTrain: Data Lifecycles and Management Plans, University of Cambridge. http://archaeologydataservice.ac.uk/learning/DataTrainDownload#section-DataTrainDownload-Module2