Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries

Research Data Management in Libraries
Workshop Report, 24-25 June 2015, LIBER2015, London

Jointly organised by LIBER, FOSTER and RLUK

Research data management (RDM) in libraries is a quickly evolving service area.

This year's workshop run by the Steering Committee on Scholarly Communication and Research Infrastructures therefore targeted two core activity areas:

1. The translation of institutional data management policies into research support services.
2. The need to support researchers in the creation and review of data management plans.

Picture: Dan North, LIBER

This report summarises the outcomes of the plenary discussions and the outbreak sessions. Recommendations on questions to consider were collected, as were practical recommendations derived from the cases under discussion. In addition, participants were asked to provide suggestions on how LIBER could help to strengthen the role of libraries in this activity area.

Day One, 24 June 2015: Translating policy into RDM support

Wolfram Horstmann, University of Göttingen, welcomed the participants and introduced to the topics of the workshop.

Andrew Cox, University of Sheffield, presented his views on data management from a researcher's perspective.

One core observation was that research is big business and competitive. It may be funded and project-oriented, but it is also very personal. A lot of work is not funded. As a consequence, researchers are often underfunded, struggle with a high administrative load, and their performance is heavily evaluated and managed by the institution and research funders.

RDM could be seen as yet another challenge to academic freedom and their independence in choosing a research subject. Moreover, the need for RDM does not seem to be really bottom-up. Funders increasingly promote an agenda of data deluge and say it should happen, and researchers need help even if they don't realise it – but: there might be a need but no demand.

This is a dangerous situation in particular with researchers, who feel under pressure. Andrew outlined a number of challenges and possible ways to overcome them: the compliance

pressure, the need for a cultural change, building empathy with the researcher's experience (librarians do research and professional development, they are not completely separated from the research process), the complexity of data and methods to uncover data (as they are not always neat things and might not be called data) and finally data sharing.

About the latter there are many concerns:

- Have I published everything I want from this data?
- Who else do I really want to share this data with?
- I wasn't funded to do this research – what's in it for me?
- I don't know where the data are, which version is quality checked?
- I haven't got time to document the data.
- What's metadata?
- My ethics application didn't mention data sharing, worthwhile to check? She has connections to companies which I'm suspicious about.
- I've got other things to do.

The main challenge here lies in dispelling the doubt – i.e. to show the researcher that data sharing is really for them.

Mike Mertens, Research Libraries UK (RLUK) presented a data typology which has been developed jointly by the Research Information Network (RIN) and RLUK. It captures and organizes terminology and is to be used by humans with the aim of facilitating data curation and discovery but also available as RDF triples (http://users.ecs.soton.ac.uk/cjg/rluk/).

Paolo Budroni and Barbara Sanchez-Solís, University of Vienna, presented the first achievements and further plans of the eInfrastructures Austria project which develops a shared infrastructure for data management and involves 20 Austrian institutions. In particular, it offers support for workflow and data management planning. Challenges in this context include the question of who owns Data Management Plans (DMPs), and who administers and reviews them.

Mijke Jetten, Radboud University, introduced us to Nienke, a 25-year-old PhD student, who works in religious studies and interviews people facing death. Here, data privacy is a major concern, but she wants to share the data and wonders how she could deal with RDM properly. There are lots of questions to solve:

- What is a safe place to store the data?
- How to work on the data with others?
- What is a safe place for sharing my data?
- Is a DMP something I can or have to do?
- Is there any training on writing a DMP?

In addition, Nienke needs to be informed about the university policy – a central policy which was introduced in 2013 by Radboud University. The policy covers storage of data and metadata, and applies to approved bachelor and master theses. The minimum storage is for 10 years. The DMP is a recommendation. The responsibility lies with the researcher, however the research director and the university jointly guarantee facilities. So far only a few researchers do as they are told by the policy.

What makes a big difference is to have a decentralised policy at the level of research institutes which, compared to the general policy, covers the why, what, how and where aspects of storing data. The library offered to help – an offer which was put on hold – and now after 1.5 years of very little progress the library is going to assist in developing decentralised policies. The RDM facilities at Radboud University are a work in progress and need to cover big and long tail data. Currently the Research Information System (CRIS) is

being adapted to allow the registration of data which can then be transmitted to one of the national repositories (DANS, 3TU.Datacentrum, etc.).

One core observation in this context is that the library is a quite stable factor. It is the link between the pilots, and it is aware of researchers' needs. With the development of decentralised policies the library is now able to transform generic services into research-specific support. For example, 75 PhD students have been trained in writing a DMP, a tailor-made course that also provides a DMP checklist. Last but not least this also resulted in the creation of possible data stewardship roles at the local level – e.g. someone who collects DMPs within his/her institution, keeps an overview of tools used for data processing, as well as an overview of storage used for data. Overall this is very exciting progress and represents new opportunities!

Sarah Jones, University of Glasgow, DCC, reported on the University of Edinburgh's approach for establishing a data policy. The policy was passed by the Senate in 2011 and created by a library-led committee which involved a range of academic champions. When released it was considered quite aspirational and its 10 high-level statements ('commandments') are quite a strong set of requirements. However, it does not come across as a mandate. A word cloud of the policy features the terms 'appropriate' and 'support' very prominently.

Responsibility is jointly shared between the principal investigators and the institution. The policy also stresses that researchers should not just hand over rights. They are encouraged to register the data and to provide project-level research data management.

Edinburgh's RDM roadmap provides a gap analysis and outlines and prioritises services. This has resulted in a rich service suite in support of research data management, ranging from the registration to the preservation of data (data infrastructures, customized DMPonline, consultancy and training sessions, etc.). Essential are the outreach activities: they visit schools and contribute to departmental meetings regularly and promote the available services.

Results from outbreak session on data policies

Data management policies from several institutions were evaluated by participant on how they provide a good starting point for developing related services. Questions discussed included which policy measures ask for further support, which forms of support are suitable, and which university unit/department should take care of these elements.

Areas that could be improved:

- Benefits for the researcher are sometimes poorly described.
- No reference to the publication of data is given.
- Deletion of data is not covered. What is the retention period before data might be destroyed?
- Legal advice is often missing. Ownership, Intellectual Property Rights (IPR) and data privacy need to be more specific. The library could perform an informing role.
- Access regulations should be mentioned (has to be specified by the group of involved researchers).
- Responsibilities: What will the institution cover?
- Long-term preservation: if the term is used, explain what it means.
- Provide references to related documents that specify further, e.g. storage policy, intellectual property policy - but don't overdo it.
- Terms & responsibilities: What does 'adequate descriptive metadata' mean? The library can give a lot of general advice, when it gets specific only the researcher can create adequate descriptive metadata.

Suggestions for LIBER:

- Offer more training events like this.
- Facilitate the sharing of practices, exchange programs, shadowing, and graduate programs.
- Provide a model policy template.
- Provide indicators on how effective a policy is.
- More emphasis on templates for specific policies.
- The library could play role in the bigger picture, e.g. training, templates, it could proactively offer the big picture.
- An example service catalogue for RDM, example implementations. Provide a short use case template.

Day 2, 25 June 2015: Support with Data Management Plans

Dan North, LIBER, presented the activities of the EU-funded FOSTER project (www.fosteropenscience.eu), in particular the current training program and the FOSTER portal. In particular, all materials are free to use, and can be edited, repurposed, and recombined. Navigation is supported by FOSTER's Open Science taxonomy. The portal also offers functionalities and guidance on how to create online courses.

Sarah Jones, University of Glasgow, Digital Curation Centre (DCC), provided a briefing on the importance of DMPs and how to create them. A data management plan is a brief plan written throughout the project to define what data will be collected, created, how it will be documented, stored, who will have access and who will be responsible. Guidance and support includes templates and often also example answers. There is some reluctance to provide boilerplate text as you want researchers to reflect but a library of successful DMPs to reuse is always useful.

More in-depth support covers online tools (e.g. DMPonline), training courses, reviewing plans and customised consultancy services. There are several areas where further developments are needed for integrating DMPs into researchers' and administrative workflows (e.g. grants costing tools).

Reviewing DMPs can be supported by guidelines for peer reviewers (compare John's Hopkins grant reviewers checklist, MRC guidelines, ESRC guidance for reviewers). To help researchers, it is recommended to focus on the benefits of DMPs rather than emphasising the compliance aspects. Don't just create a plan for compliance and not look at it: use it as a tool and a guide during the research, and update it where necessary. Currently, most research funders are not checking at all, often not even on a basic level, good enough to be funded – but this is about to change.

Suggestions for LIBER:

- Pull more DMP examples together, internationally.
- Promote towards research funders to handle costs for data management and data infrastructure as a separate budget compared to the research funding budget.
- Collect evidence on how many DMPs are actually lived, i.e. statistics.

Mari Elisa Kuusniemi, University of Helsinki, talked about reviewing DMPs for Horizon 2020 (H2020) Grant Proposals and how it all started when they prepared themselves for the data management services.

The University of Helsinki currently has about 40,000 students, with a library system of about 240 staff members across 4 campuses, and its research data policy was published in February 2015. However, the service preparations started much earlier, with in-house training since 2012, and a DMP education in January 2015 by expert from Finnish Social Sciences Data Archive.

This also involves networking with other service providers at the University of Helsinki (e.g. lawyers, researcher funding services, the advisory board on research integrity, etc). Moreover, there is collaboration with other service providers in Finland (e.g. the CSC which runs a central data infrastructure and services).

In preparing workshops for the library's research data team, all H2020 guidance was explored and the following themes were covered:

- Why DMP consulting is so intimidating?
- How can we improve our expertise of DMPs by self-learning?
- How should we team up to survive H2020 reviewing?
- What does it mean to work as a team member?
- How do we deliver the DMP service?

In most cases the research funding services asks the H2020 projects to send DMPs to the library for reviewing and consultation. To answer questions, the LibAnswers-tool as service desk is used. The DMP consultancy team consists of 5 people (RDM is 20-50% of their job).

<u>How do we do it?</u>

First of all, the DMPs are read, and a meeting with DMP consultancy team is offered. Help from other services providers is requested if needed. In general, researchers are asked to further define the DMP – they are often very difficult to read. Example questions can include:

- If the empirical data is collected in Italy, where will the data be collected and processed at this stage?
- How is your data protected during the project?
- What does this abbreviation mean?

Future plans will target improvements of the process, better guidance, and a DMP tool of some kind (a Finnish project is already on the way), and to get all Liaison librarians involved.

A core lesson learned is that support for research data certainly needs to be based on a multi-professional team: IT experts, lawyers, and of course data management experts – from a personal perspective you will need courage and attitude.

<u>Practical suggestions from the discussion:</u>

- How do you prepare yourself? A lot of networking with stakeholders, contacting them directly and scheduling meetings, asking them to join network meetings (3 times/year).
- What is the role of the library in your data management? The library has the main role for DMPs, with the help of others, e.g. IT – they will appreciate it as they know that they do not have too much work with researchers if they work with us.
- A Roadmap, the library is responsible for DMPs, basic education, finding data repositories, finding appropriate data repositories, etc.
- Staffing at the University of Helsinki is based on a team of five, comparatively very high compared to the UK. Of these three persons to educate the other 9, e.g. liaison librarians, very important to know the subject area, quite fast to be ready to provide

advice as a service. The service itself is very demanding service, the researchers asking us more and more. This also raises the need tools to become more effective.

- Budgeting in project: We promote to add a budget for curation and infrastructure, and the documentation of data. But so far we have seen only very few examples yet, e.g. biomedicine, but not in the social sciences and humanities.
- Any role for the library in the projects? They are asking us every day, but no resources. The library helps with metadata standards, the selection of the standard but not with using it. Currently we only have few experts who have to educate people, with more resources this task could be covered as well.

Results from the outbreak session on reviewing DMPs

Participants reviewed data management plans (DMPs) in groups by research area: humanities, social sciences, sciences. Questions to be considered included what data and how it is collected, how it is stored and shared, what needs clarification and refinement. **NSF Grant DMP**. The plan mentions storing data for video films, software, computer-generated data, written from IT security point of view. It didn't say why they wanted this data repository: add a description of the roles and responsibilities (unclear of who is performing which tasks). The text is not structured very well. It doesn't discuss the costs of storing video files, for example. How should data be referred to (DOIs, etc.?). They say the data would be shared in a 'public space', but what does this mean? They say how the work will be disseminated (conferences etc.), but what about copyright? No mention of metadata. In addition, it needs a more systematic approach to storage rather than sending it in email.

DMP Radboud, Law department. What is the help of the library when reviving this plan, help in archiving the data – write down where to get this. Describing data during research: there is a mismatch on where to put the files, unsure about how to deal with the data, how to analyse the data. Issues include: save files on own laptop, not save, could be stolen - advise her to work on a networked drive, box-like account, and make sure that the data is safe. Provide a document that is helpful to her, in how to store, organise her files. Regularly backed up, refer to standards. Ethical committee: how to get consent, e.g. somewhere abroad (Argentina in this case), make clear if such a committee is not available. What is the role of the library? Could they help more in archiving and storing? There seem to be issues about describing the data during research, a mismatch between which files to store where: researcher seemed unsure about dealing with data (she is from a law background). She said she would be saving files on her own laptop, which is not a secure storage medium. Advised to work on network drives. Questions in DMP over the ownership of data remain unclear. Researcher says she will update the DMP when she can answer that question.

Arts and Humanities. The project will be creating videos about rituals and festivals. Management of data and sharing seems to be quite straightforward, lots of details on formats. Not much about ethics and legal issues. She had asked for participants' consent on sharing the videos, but had no back-up plan for when they refused consent. There were issues around citation; which is the authoritative copy of the video file if it is available in several different places (UK data archive, Vimeo).

Economics. The project creates a portal about tax information. It is a partnership between two universities and commercial partners, so there were concerns about costs and how these would be shared. More information about metadata, description, cataloguing is needed. They want to make available information that is currently held by proprietary institutions.

Climate change. Very important when reviewing is to talk it over with the researcher, s/he has to be there. The plan was sometimes very strong but in other places it lacked information. Long-term preservation remains unclear, how to analyse the data, raw data, how to cite the data. Before the researcher start writing think about who is going to read it.

Chemistry. The DMP is quite well-written but many things need to be clarified, e.g. formats, responsibilities, and others are missing, e.g. information about licensing and costs. Some paragraphs are just standard phrases, and some are in the wrong place. The plan needs clarification on document formats; what kinds of metadata are they referring to; needs to clarify who is responsible for collecting data and other roles. No mention of licensing and costs.

Picture: Dan North, LIBER

Economics RC ESRC. The plan starts off very strongly, describes what data is going to be produced and analyzed. However, it ends off quite vague and it is not clear who has overall responsibility across involved partners is. Regards IPR it is not clear whom to contact about it. There is no designated contact about archival questions. The plan remains vague about sharing among partners (e.g. requests via e-mail). Some information will be provided through institutions websites. There are no figures on costs. UK institutions servers for storage but no details are provided. No procedures if any difficulties occur, for example with sharing. No procedure for sensitive data.

<u>Suggestions for LIBER:</u>

- A role for the library: LIBER could define what the library should do, and define the roles for the plan. Libraries are an ideal place to locate a network, advocate for this in a European context, to promote the role for librarians in RDM. Change the culture in libraries to where data will eventually be on an equal footing with publications.
- Budgets: A heated discussion. LIBER could help to inspire regulations, provide arguments on how to argue for budgets on a European scale, how to collaborate to make it cheaper, etc.
- Education & training: Libraries should be influencing library school curricula, and LIBER could offer accredited courses.
- Reviewing DMPs: Putting yourself in the shoes of the reviewer and going through a few DMPs, helping them to make a better DMP. LIBER could provide advice on how to clarify questions about the ownership of the data. More sharing of experience of reviewing DMPs. Overview of funders' expectations. Role of the library: only during grant proposal, or also at a later stage.
- Tools: LIBER could create an international library of DMPs, a metadata standard for harvesting DMPs, advice on sharing DMPs. More researcher use cases to show benefits for researchers, e.g. videos. A library of use cases, include good AND bad practice examples.