Activities, solutions and experiences within UK universities to meet the EPSRC research data policy

Monica Duke, DCC, Jonathan Rans, DCC
Verena Weigert, Jisc
University of East London

Interview with Stephen Grace, Research services librarian

19 January 2015

Interview led and recorded by Verena Weigert, Jisc and Monica Duke, DCC

Image courtesy of Stephen Grace - https://datamanagementuel.files.wordpress.com/2013/07/20130626_094624.jpg
Introduction to the University of East London (UEL)

UEL is based on two campuses in the heart of East London. It has an ambition and aspiration to be the ‘leading modern, post-92 university for research’. There are seven disciplinary schools. UEL received support from the DCC in an institutional engagement during 2012-13. Stephen Grace was appointed research services librarian in October 2011.

Policy, strategy, governance and sustainability

Policy development

In the last quarter of 2011, UEL took the step of recruiting a research services librarian with a specific remit to improve RDM. One of the first tasks that he undertook was the writing and adoption of an RDM policy for UEL. Working as the lead for a small project team, a draft policy was developed, inspired by policies existing at the time within other universities. Feedback on drafts was obtained from the DCC and the project team, particularly about the language to use and areas to cover. Research leaders in each school also gave feedback. The turnaround was quite quick, boosted by the agile decision-making culture at UEL, and the policy was formally adopted in March 2012.

UEL chose a policy-first approach, trusting that the policy would represent a concrete and visible commitment from the institution to allocate funding internally for services and infrastructure. The library service was named in the policy as the party to develop the RDM services (in consultation with departments and researchers) and it continues to provide leadership for RDM to the institution. UEL chose not to appoint an RDM steering group to oversee implementation of the policy, instead its research committee provides general oversight for RDM. The Research Committee is representative and includes senior academics and representatives from services (IT, the library and research and development support). In practice, the committee has entrusted the research services librarian with a remit to move things forward without requiring approval for each decision, for example a programme for staff/student RDM training was designed and delivered without the need to go to the committee for approval. The committee remains available for consultation when major decisions are needed that would affect schools.

2 A full list of policies can be found on the DCC website: dcc.ac.uk/resources/policy-and-legal/institutional-data-policies. It should be noted that at the time the UEL policy was developed only a small handful of examples were available, making the UEL policy the fifth to be listed on the DCC website.
3 UEL RDM policy: http://dx.doi.org/10.15123/PUB.3539
4 In early discussions about institutional policy, an alternative view emerged, that a policy that was approved without the infrastructure and services in place to deliver the policy implementation, would only set up expectations that could not be met, possibly leading to disappointment or disengagement with the RDM agenda. This did not prove to be the experience at UEL.
Resourcing

Like many institutions, UEL was galvanised by the EPSRC requirements but it also had an existing intention to strengthen the research base. Library staff responsible for RDM services keep in contact with academic researchers to understand their needs.

Alongside the research services librarian, who is now well established in post (but has other responsibilities in addition to RDM), UEL has also funded a full time RDM officer to assist.

Other central funding has been committed to develop a data repository and a case was made for storage of resources that the library itself creates. An archival server has been set up for all data resources and there are plans to join it up with the repository.

In terms of ongoing costs for active data storage, PIs usually contact IT services directly and have their needs met that way. There is no recharging model internally, therefore a costing model for data has not been required. Monitoring and discussion of this situation continues in conjunction with the director of IT services. Whilst a survey of researchers did reveal some demand for storage information, most were requesting training and guidance for themselves and students (including knowledge of compliance requirements) and this is where the priority for the RDM service was focused.

From policy to service

Although UEL does not preclude providing oversight for RDM development through a steering group in the future, this is not felt necessary at the time of writing. A roadmap (as requested by EPSRC) was in place by 2012-13.

The RDM service, called research data services (RDS), is run as a one-stop shop. When requests require specialist input, RDS refers academics on to the specialist sections of the university that they work with closely, namely the legal office for intellectual property rights (IPR), ethical and legal issues, the finance office for costings, or ReDS (the research office) for input into bids.

One of the main services offered is help with drafting of data management plans (DMPs). DMPonline is used to produce a draft of the plan, with input from academics. The academics then take the draft and complete it, with ongoing discussions with RDS, when needed. UEL has produced its own templates for DMPonline, including one for students and others for when funders do not have a specific requirement for plans (e.g. the NHS). Although the need for the EPSRC template has not yet arisen, the potential to provide plans for EPSRC bids was a big driver to mitigate against any disadvantage in being able to bid for EPSRC in the future. Part of the advice given by RDS is to stress that funding councils expect costs for time dedicated to RDM to be included in bids, for example allowing time at the beginning of projects to write protocols for dealing with data and at the end, to prepare data for archiving.

Overall, institutional policy development is viewed as a quick win at UEL. With the policy in place and commitment from the institution, the argument could be made for allocating funding internally for services. The institution is quite agile in its decision-making and in allocating resources. There is a culture of engagement with central services by academics. Services like the library are held in high esteem; they are seen as trustworthy partners from the outset, which helps to make things happen.

DMPonline: [https://dmponline.dcc.ac.uk](https://dmponline.dcc.ac.uk)
Data management support and staff development

Providing training

A key strategy for training and awareness to meet EPSRC expectation I at UEL is to work closely with the graduate services on their development programme. UEL was proactive in securing funding for a Jisc project on training which ran between July 2012 and May 2013. As part of TRAD some workshops were developed and offered to academic staff and graduates. The workshops were very well received and have continued to develop since the end of the project. They are now offered regularly through the graduate programme as four different two-hour workshops, on the themes of ‘General introduction to RDM’, ‘Good practice in RDM’, ‘Writing your DMP’, ‘Sharing and archiving data’, and a fifth workshop is in development.

Besides the generic training, a bespoke course for the clinical psychology doctoral programme was delivered. This is now delivered annually and receives good feedback. Lecturers are informed about the training that their doctoral students are being offered. Another benefit of offering training is that follow-up contact often happens, as in the case of a separate school that was also offered doctoral training.

On-demand training is available in response to requests from academics. One school research group wanted to introduce mandatory DMPs for students as a planning tool when working with their supervisors and this idea is now being considered by other schools. Discipline-specific training can be developed but does require modification to tailor the content and examples to the discipline. Training on DMPs, in particular, requires close working between RDS and academics themselves who have to provide examples of DMPs.

Making contacts and links

UEL raises awareness of RDM requirements and promotes good practice by offering support and proactively developing relationships. The RDM team goes out to meet people in their departments and takes care to meet their stated requirements specifically. Other means of contact are available such as a mailbox for RDM services although, due to the small size of the institution, the RDM staff are known to the academics and often establish contact through their own personal email addresses. In practice, a small volume of requests comes through the mailbox.

One other route to receive requests for help is by referral from the research office, which explains the need for DMPs to researchers, signposting the library as a place where assistance is available. The RDM staff are able to engage one to one with those writing DMPs. A few people in the institution had already written plans when the RDM service started and the training in this area was considered timely and very welcome. Training sessions are also valuable in providing direct contact with academics. The website that offers RDM advice at UEL is in the process of being expanded, following web infrastructure changes at the institution.

Other materials produced by TRAD included an online module (supportDM) aimed at (and tested with) subject librarians, which gave an overview of RDM. At the time it was envisaged that subject librarians might take on

6 TRAD project website: uel.ac.uk/trad
7 UEL RDM webpages: uel.ac.uk/researchdata/
8 Available in Jorum for reuse: jorum.ac.uk/
roles in RDM, although the recruitment of a dedicated post superseded that need. These materials may in future be re-used with staff from IT services.

UEL has also trialled the use of drop-ins for RDM support. These had mixed results because publicising them was challenging, and turnout was low; however, the contacts made were rich and resulted in useful in-depth conversations. This type of outreach is currently on hold: it is seen as having value, but work is needed to find an effective way to promote it.

Throughout the process of offering support and training, the EPSRC requirements were not positioned as the focus. However, the similarity of requirements between councils means that the required training needs were met.

Infrastructure

A data repository based on EPrints

UEL runs two repositories based on the EPrints software, one for publications and one for data. Data.uel.ac.uk has become recently available, and contains a small number of datasets (including statistics on the use of the publications repository). Software developed at the University of London Computing Centre (ULCC) will be used to make explicit links between datasets and publications. One dataset that has been published offers a good example of what will be possible. This is a dataset on public health, consisting of a major survey in two Indian states, with comprehensive documentation alongside the SPSS dataset. This data collection links to two publications held by the UEL publication repository, two publications held in external repositories and baseline data held elsewhere. The example will be used in outreach activities with other researchers.

The software used is an adapted version of the ReCollect EPrints plugin that facilitates linking. In practice it was found that an out of the box solution was not available for making data available openly.

A new IT helpdesk ticketing software system will be employed to help record requests for access as required by the EPSRC expectations. Requests for data that is only being made available on application will be recorded, and sent on for approval to the relevant PI or designated data steward.

The separation between the publications and data repository reflects different policies and workflows. The publications repository only holds open access (OA) items, whilst the data repository may have some restricted access. It was felt that it was best not to dilute the message about OA that is associated with the publications repository. Different levels of checks and mediation are applied to the two repositories, for example, only a limited OA for the data can be carried out by the library (e.g. basic checks for personal data, supporting documentation) and the knowledge for describing the data must come from those who created it. ULCC developed a special linking mechanism between the two repositories to relate data collections to publications; this is now available.

9 EPrints website: eprints.org
10 Published at: http://dx.doi.org/10.15123/DATA.4
11 ReCollect EPrints plugin: http://bazaar.eprints.org/367/
12 Repository links: http://bazaar.eprints.org/379/
Activities, solutions and experiences within UK universities to meet the EPSRC research data policy
Monica Duke, DCC, Jonathan Rans, DCC
Verena Weigert, Jisc

Identifying the need for support

The UEL RDM policy makes an assumption that data will be shared wherever possible, with the onus being left to the PI to offer up the data to be managed by the institutional repository. The DCC guides will be used to facilitate the discussion with the PI about preparing data for archiving, the value of the data and processes like anonymisation and also to answer questions around documentation.

Links with the research office are considered vital in determining which projects will need help and support with RDM, so that this can be offered. The granting of an EPSRC grant should trigger notification from the research office to the RDM staff so that they can reach out to grant holders.

In terms of reviewing how long data should be held, there is a mechanism to check every five years. A flag for the review date is included in the repository, and a manual intervention to evaluate retention will be made. In essence, the ingest step for data is where the major effort lies, and retaining data following ingest is considered a low-cost activity.

Non-digital materials will require more attention in future. Generally, physical items (such as samples) are looked after by specific departments, although the institution offers a small archive for paper-based materials. Other physical objects from funded research activity could come from the fine arts school.

Identifiers and metadata

UEL has signed up with DataCite to be able to assign DOIs. All items deposited in the data repository are assigned a DOI\(^ {13}\). This is done in a controlled manner, once an item has been deposited a mediated service makes the data ‘published’ and the DOI is assigned. The plugin used for EPrints was developed by ULCC and it has been modified to enable the interactions with the DOI service. The plugin is available to the community\(^ {14}\). Control of the suffix portion of the DOI is also possible.

UEL has opted for just the five mandatory elements of metadata recommended by DataCite. Although the ReCollect plugin is used the metadata elements specified by it were considered too many. The deposit workflow is split into two screens so that the essential elements are displayed in the first screen, with a second screen showing the optional elements. The service relies on academics with the specific disciplinary knowledge to supply the metadata.

The repository also acts as a registry for data that is held externally, for example, some ESRC-funded data will be required by the funder to be deposited in RESHARE. Metadata about non-digital items could also be recorded in the repository. The development of the plugin has been shared with the EPrints community at the user group, with a good response and it is expected that other users of EPrints could also benefit from using the plugin.

These arrangements will position UEL to be ready to meet EPSRC expectation V, and should mean it can accommodate expectation VI.

---

\(^ {13}\) Active consideration is being given to assigning DOIs to theses.
\(^ {14}\) DataCite DOIs for EPrints: [http://bazaar.eprints.org/380/](http://bazaar.eprints.org/380/)
Challenges and top tips

Reflecting on its experience, the following thoughts and tips from UEL can be passed on to others:

» Lots of material available for re-use, from the DCC and from the outputs of the Jisc MRD programme, now deposited in JORUM – finding material to re-use for training is a quick win

» Finding discipline-specific material is more of a challenge but there is scope for sharing more of these materials between institutions (and there may be a role for Jisc here)

» When meeting researchers, it pays to do some work beforehand to find out their main queries

» Links with other service departments are very important. For the UEL RDM service, the relationship with the graduate school has been a key way to input into training. The service also relies on intelligence from the research office about new grant holders who may need support, as well as direct referrals

» The disciplinary knowledge of data creators will be relied on for description of datasets during the deposit process, as a self-deposit model will be implemented (with some support for deposit in place)

» Dealing with non-digital items is an ongoing challenge – finding physical storage space, deciding how to archive it and determining value all need to be considered

» The processes of getting a policy in place and running workshops were the most straightforward and painless. Putting a policy in place should be another quick win, especially now that several have been produced, published and are available for reuse

» One of the major successes was working with the graduate school to run training. Evaluation showed that the training worked well and helped to build trust between the services

» The support of the DCC was valuable as a sounding board and for feedback, especially as RDM professionals are working alone or in small teams

» A particular challenge has been timing the inclusion of IT services in the planning for an RDM infrastructure. Prior commitments to support the Olympics (the UEL campus was used as a telecommunications base) followed by a major programme of renewal of the IT infrastructure mean that discussions with IT were put on hold while other areas were progressed

» An online survey of research needs was designed based on the data asset framework (DAF)\(^\text{15}\). The results of the survey exemplified the practice and the issues that people had and confirmed that the approach being taken for an RDM service was right. It highlighted problems that had been encountered (e.g. lost data), identified the need for training, and also showed the need to tackle RDM across different services (which has been useful to other service leaders such as IT)

» Training has been a great way to build working links, the training sessions have been rewarding not just in and of themselves. They have resulted in lots of follow-up with staff coming back to ask for help.

\(^{15}\) DCC Data asset framework: dcc.ac.uk/resources/tools/data-asset-framework