Key issue

A UK Research Data Service (UKRDS): the way forward for research data management?

JOHN MILNER
Project manager
UKRDS

The background

In many research disciplines there is a growing deluge of digital data. Capacity planning estimates from the University of Bristol, for example, indicate storage demand growing at 127% per annum. This is not a problem that will go away. However, it presents opportunities as well as challenges. Clifford Lynch, Executive Director of the influential Coalition of Networked Information (CNI), recognized this in his article 'Big data: How do your data grow?' (*Nature*, 4 September 2008) when he said:

> *"Because digital data are so easily shared and replicated and so recombinable, they present tremendous reuse opportunities, accelerating investigations already under way and taking advantage of past investments in science"*

This issue and opportunity is not restricted to science either.

Consequently, the Russell Group IT Directors (RUGIT) and Research Libraries UK (RLUK) proposed a HEFCE/JISC-funded study into the feasibility of a shared national service for research data management (UKRDS)[1]. This study began in March 2008 and its report and recommendations went to HEFCE in December 2008.

The findings

The study worked closely with four 'case-study' universities (Bristol, Leeds, Leicester and Oxford) and a large body of stakeholders including major existing service providers. From this work it was concluded that, while there are pockets of good practice, and some excellent services exist, these facilities are not 'joined up'. Provision for researchers therefore lacks coherence and there are many gaps in provision for research data management.

The study team used questionnaires and workshops to try to understand researchers' needs for data management. Through this work we were able to consult approximately 700 researchers.

In looking at the retention value of data, it turns out that about 50% of data is estimated to have a useful life of less than ten years; 26% is seen as having indefinite retention value. In terms of sharing, most research data is held locally, with 21% of questionnaire respondents using a national or international facility.

Most researchers share data but mainly within research network teams and collaborators; 18% share via a data centre and 43% would like access to others' data.

Different attitudes amongst researchers towards the potential of a UKRDS often boiled down to the presence or not of a national data centre for their subject.

The study also highlighted the limited levels of data management skills amongst researchers and the consequent need for the further development of training capabilities.

It became clear that for a shared service to succeed there would be a need for a coherent approach to data management, with practical procedures and standards and the necessary support and training for their introduction. Moreover it

would need to be an 'opt in' service for both researchers and existing service providers. This is because not all research data needs management and, where it does, some of the requirement is already well addressed through services operated by ESRC, NERC and others. Those existing facilities are strong and stable and would need little from a UKRDS. They may, however, be able to add value from development of their facilities in support of a UKRDS.

The work of the Research Information Network (RIN)[2] and Digital Curation Centre (DCC)[3] has also proved to be invaluable and the DCC life cycle model in particular has informed the recommendations that follow.

The study team also investigated the international perspective and found that much work in this area is being conducted in the USA, Canada, Germany and Australia, where a similar shared service, the Australian National Data Service (ANDS)[4] is already under development.

The conclusions and recommendations

The study team concluded that a national shared service such as UKRDS was certainly feasible, though challenging to implement.

The key to engagement with the proposed service is the concept of a data management plan (DMP). The purpose of the DMP would be to carry the policy for metadata together with other essential information to determine how a data set would be managed through the data life cycle defined by the DCC Data Life Cycle Model[5].

The team believes that it would be for researchers and funders to determine if the data set(s) produced by a research project were suitable for medium- or long-term management. If so then (unless there is an existing subject-based facility) the research group would expect their funding to include resources to work with UKRDS to produce and implement a DMP.

A UKRDS would build a registry of such DMPs with two purposes in mind:

1. To provide the framework for the management of data sets after the projects that produced them were completed.
2. To provide the core of a data discovery service for the global research community, by making the registry searchable.

At this juncture it is important to point out that a UKRDS would not be about the *storage* of data but about the *management* of data, which could be stored in any suitable facility as determined by the DMP.

The benefits of a UKRDS include:

- protection of investment in research and extraction of greater value
- preservation of opportunities for future research
- informing the strategic development of the research infrastructure
- reduction of duplication, recreation and errors, and unplanned data loss
- volume growth/capacity planning is more cost-effective
- more opportunity for re-use, cross reference and data-set integration
- better targeted retention and disposal
- shared skills giving better coverage and thus better productivity in both service providers and researchers
- proper focus for practical best practice.

Additional direct benefits to an HE institution, to the researcher and to funders, could include:

- providing a focus for promoting the work of the institution and the researcher
- guidance on which repository to get research data from and a gateway to trusted service providers
- commissioning new services to fill gaps in data management.

In addressing the approach to implementation of a UKRDS it was clear that some sort of phased approach would be needed to make the task manageable. However, it was also clear that most of the core service elements of a UKRDS would need to be in place to manage even a small number of data sets.

The study team therefore defined the concept of a Pathfinder service. This would allow the development of a complete service with a small number of projects, mainly from the case-study universities, working with a restricted set of service providers. The dual funding model would imply that research councils would be called upon to fund the project-related work on DMP while the funding councils would support the development and establishment of the UKRDS infrastructure. Once established, the Pathfinder could be scaled up, driven by demand and constrained by resources,

until it was serving the whole target HE research community.

Should the recommendations in the report to HEFCE be accepted, then we hope that funding will become available to allow the start of the Pathfinder development during 2009.

References

1. UK Research Data Service (UKRDS):
http://www.ukrds.ac.uk (Accessed 9 January 2009)

2. Research Information Network (RIN):
http://www.rin.ac.uk/ (Accessed 9 January 2009)

3. Digital Curation Centre (DCC):
http://www.dcc.ac.uk (Accessed 9 January 2009)

4. Australian National Data Service (ANDS):
http://ands.org.au/ (Accessed 9 January 2009)

5. http://www.dcc.ac.uk/docs/publications/
DCCLifecycle.pdf

Key issue © John Milner

To view the original copy of this Key Issue, published in *Serials*, the journal of the UKSG, click here:

http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=22&issue=13&spage=83

The DOI for this article is 10.1629/2283. Click here to access via DOI:

http://dx.doi.org/10.1629/2283

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=22&issue=1