

Meeting the Research Data Challenge Managing Research Data programme 2009-11

Briefing paper

November 2009

Researchers in almost all disciplines now create data in digital form. These data can come in many guises: for example, the measurements recorded by environmental monitoring satellites, the products of collisions between fundamental particles, the sequences of entire genomes, the results of social science surveys and interviews, the annotated images of ancient Greek inscriptions or the annotated videos of innovative dance routines.

Like the books in a well-run library, some of these data are curated and kept for future access and re-use in well-managed data centres that are usually subject-based. These lucky data, which typically are generated by large-scale facilities or major goal-oriented research programmes, can be re-analysed or interpreted by tomorrow's researchers who may use them to answer questions we cannot predict today. The majority of research data, however, goes uncatalogued and is therefore not reusable. In all probability, this is not only a loss to posterity, but also a failure to reap the full potential from present day investment in research.

As the volume of research data increases relentlessly, the need to find ways to manage and curate data for sharing and re-use also grows. Many organisations have a stake in finding solutions, which will vary almost as much as the data themselves. Policies are needed at international, national and institutional levels that are rooted in the actions researchers themselves will need to take.

Research data is therefore a complex issue and devising policies and solutions is no simple matter. Through its newly launched Managing Research Data programme, JISC is plugging gaps in our knowledge and illuminating some of the key outstanding questions, for example:

- Are our assumptions about the value of research data right?
- What do researchers need to manage data effectively?
- What policies are required to support researchers' requirements?

Researchers, data managers, librarians, institutional managers, funders, policy makers and government all have a stake in the answers.

Helping researchers and institutions

Institutions are coming under increasing pressure to manage the research data generated by their researchers that cannot be curated by subject-based data centres – and

many are unsure how to proceed given the absence of clear good practice. To address their concerns, JISC's Managing Research Data programme is funding projects to provide the UK higher education sector with examples of good research data management.

The projects are first identifying requirements to manage data created by researchers within an institution, or across a group of institutions, and then piloting research data management infrastructures at institutional, departmental or research group level, to address these requirements. Cost-benefit analyses are included in the work. The projects' experiences of using some existing tools to help with planning these infrastructures will be used to develop and refine the tools, which include the Data Audit Framework (DAF) and Assessing Institutional Digital Assets methodology.

Improving the way in which research data is managed will rely on close partnership between stakeholders and an understanding of the impact of disciplinary priorities, existing data management provision, funders' policies and institutional context. JISC is contributing to this understanding by exploring best practice in data management planning across a number of projects funded by the UK research councils. By working across a range of subject areas and funders, this initiative will furnish detailed case studies, casting light on data management issues, illustrating the range of problems and their solutions and providing model data management plans.

What the Managing Research Data programme will do

- Develop exemplars of good research data management
- Build on and refine existing tools for research data management
- Capture best data management in research council projects
- Data management training developed and delivered to postgraduates, librarians and research support staff
- Explore data publication in journals
- Improve tools for citing, integrating and linking data
- Assess the value of curated and uncurated data
- Draw up a research data roadmap

Developing capacity and skills

Effective data curation requires skills associated with information managers, 'informaticians' or librarians in combination with the detailed knowledge of subject area specialists. There is currently no accepted and professionally esteemed route to acquiring these skills and few people therefore have them. Bridging this skills gap is perhaps the most significant challenge of all which, some argue, will only be met by targeting subject-specific, post-graduate research skills training.

JISC has been working to address this challenge for some years by supporting the Digital Curation Centre (DCC) which has developed a set of training materials, including the much applauded Digital Curation 101 course. The Managing Research Data programme is building on this work by supporting the development and implementation of data management training, based on the DCC's experience, for postgraduates, librarians and research support staff.

Publishing, citing, integrating and linking research data

Researchers need incentives to curate their data and make it available for re-use. This could be provided by publication in a scholarly, peer-reviewed, online 'overlay' journal that champions the sharing of data as an output of research activity. Such a publication could provide a means of locating and accessing datasets and of drawing attention to well-curated research outputs in institutional data repositories. Data could then be cited in their own right, be linked with other data, resources, articles and people and integrated across disparate sources. However, conventions would need to be agreed for data citation and protocols for automatic linking.

The Managing Research Data programme is exploring this vision by funding projects to pilot models for data centric journals and to explore and improve the conventions, tools and techniques required to cite, integrate and link research data.

Developing a strategy for research data

Many agencies and researchers already have considerable experience of research data management and this expertise must be integrated and built upon, whether it is provided by subject-based repositories, national and international data

centres and archives or higher education institutions. Such integration, however, will come at a cost and we need to be sure that uncurated, unreusable research data does indeed represent a significant untapped resource and the investment will therefore be worthwhile.

A study, funded jointly with the Research Information Network, is measuring the use and impact of several UK-based research data centres in different disciplines, including the Archaeology Data Service, the National Cancer Research Initiative/Information Network, and the UK Solar System Data Centre. This will be followed by an examination of the value, benefits and opportunity costs represented by institutionally held data sets, curated and uncurated, using existing findings including those generated by some of the projects under the Managing Research Data programme mentioned above.

Assuming there is a strong case for data curation and preservation, many questions follow. For example:

- How can we increase the capacity for managing and curating research data?
- Who should take responsibility for pursuing such a programme and who pays?
- If the solution, as seems probable, involves a complex mixed economy, with data being curated by national data centres, HE institutions and collaborations across institutions, how is such a system to be created and organised?
- Under what circumstances will institutions be responsible for data management?
- How can we best leverage the expertise of subject specific data centres?
- What blend of skills and roles and what level of national coordination is likely to achieve the optimal solution?

These relatively open questions are subject to intense debate. The Managing Research Data programme is establishing an advisory group to draw up a Research Data Roadmap towards a coordinated national solution. The group is being constituted through a Memorandum of Understanding between JISC, HEFCE and the research councils to make it representative and give it appropriate authority.

Written by Simon Hodson, JISC Programme Manager for the Managing Research Data Programme.

Alternative formats of this briefing paper can be found at:
www.jisc.ac.uk/publications

Further Information and Reading

The Managing Research Data Programme
www.jisc.ac.uk/whatwedo/programmes/mrd

The Skills, Role and Career Structure of Data Scientists and Curators: an Assessment of Current Practice and Future Needs
www.jisc.ac.uk/publications/documents/dataskillscareersfinalreport

To Share or not to Share: Publication and Quality Assurance of Research Data Outputs
www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs

The UK research data service feasibility study: Report and Recommendations to HEFCE
www.ukrds.ac.uk/resources

Digital Curation Centre
www.dcc.ac.uk

Data Audit Framework
www.daf.ac.uk