



**DATUM for Health
Research data management training for health studies**

Final Project Report : Project Summary

Prof Julie McLeod, Northumbria University

August 2011

Funded by JISC, under the Managing Research Data (JISCMRD) Programme

This work by Authors:

Julie McLeod, Northumbria University, School of Computing, Engineering & Information Science

Project Lead

Professor Julie McLeod, Northumbria University, School of Computing, Engineering & Information Sciences

Project Team

Professor Charlotte Clarke, Northumbria University, School of Health, Community and Education Studies

Professor John Dean, Northumbria University, The Graduate School

Kevin Ashley, Director, Digital Curation Centre (DCC) <http://www.dcc.ac.uk/>

William Kilbride, Executive Director, Digital Preservation Coalition (DPC)
<http://www.dpconline.org/>

Sue Childs, Northumbria University, School of Computing, Engineering & Information Sciences

Elizabeth Lomas, Northumbria University, School of Computing, Engineering & Information Sciences

Project websites

Northumbria University: <http://www.northumbria.ac.uk/datum>

JISC: <http://www.jisc.ac.uk/>

JISCMRD Programme: <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>

Acknowledgements

The team would particularly like to acknowledge the contributions of:

- Ms Laura Smiles, The Graduate School, Northumbria University for all her support and enthusiasm in organising, promoting and participating in the training programme
- The Advisory Panel Members from the School of Health, Community and Education Studies; Research Business & Innovation; University Library and Learning Services; and the University Secretary's Office
- The Postgraduate students who participated in the programme and gave their evaluation freely
- Co-organiser Ms Carol Jackson, DPC and presenters at the Data4Life event (<http://www.dpconline.org/events/details/28-data4life?xref=27>)

This report is made freely available under a Creative Commons Attribution-NonCommercial-ShareAlike 2.0 UK: England & Wales License.

© Copyright holder: Northumbria University

Project Summary

Why we did the project

Good research data management (RDM) is required wherever research takes place, and the importance of managing research data effectively and efficiently is being emphasised by national and international external drivers. Three key drivers are: research funder requirements¹; regulatory compliance; the demand for greater openness, transparency and accountability. Additionally, developments in digital technologies have changed the way research is done, providing new opportunities for doing research and also presenting challenges for managing research data and records. However, infrastructure (i.e. policy and procedures, human and technical) to support RDM is underdeveloped and many stakeholders in HEIs (PGR students, academics, researchers, research support staff) lack knowledge of RDM and the necessary skills to implement it. To promote RDM skills the JISC-funded five projects “to design and pilot the delivery of course materials that HEIs can use to improve the skills and abilities of academics, academic support staff and postgraduate researchers” as part of their Managing Research Data programme.² *DATUM for Health* was one of those projects.

What we did

DATUM for Health was a collaborative project between Northumbria University, the Digital Curation Centre and the Digital Preservation Coalition. It focussed on postgraduate research (i.e. doctoral) students in the health studies discipline and on the management of qualitative, unstructured data³. The training programme comprised four sessions, delivered over a period of weeks, and covered both generic and discipline-specific issues.

- Session 1: Introduction to research data management
- Session 2: Digital curation 101 lite
- Session 3: Problems and practical strategies and solutions
- Session 4: Data4Life - Digital preservation for health studies

Northumbria University delivered sessions 1 and 3; the DCC delivered a tailored session on digital curation (session 2), and the DPC jointly hosted a roadshow session on digital preservation in health (session 4) which was opened up to a wider audience⁴. The programme was piloted with postgraduate research (PGR) students in health studies who were full-time / part-time and at different stages in their doctoral study. 25 people took some component of the programme (22 students, including one from outside the School of Health, Community & Education Studies and two from other regional universities, plus 3 researchers).

What we learnt

About PGR RDM training: In their evaluation participants rated the pilot programme very highly, finding it useful and professionally run. Whilst holistic in its design, participants who missed some sessions did not feel disadvantaged and felt the sessions could also ‘stand alone’. They preferred the face-to-face delivery (a mix of lectures, exercises, group work, discussion, directed learning) and discipline-specific training but most were not concerned about methodology-specific training (i.e. covering only qualitative or quantitative data). Only 11 students joined the *DATUM for Health* organisational site on the VLE mirroring the experience of The Graduate School, with PGR students not being big users of the VLE. Participants thought RDM specific sessions were necessary, though RDM could also be covered in other courses in the PGR training programme, e.g. induction. Whilst they enjoyed the full day event (final session 4), which gave a wider perspective of RDM from a range of practitioners, they felt three (2.5 hour) sessions would be sufficient in the future. These should be early in the PhD journey (e.g. before Project Approval) though RDM training would be useful at other stages too. They found RDM, and the DMP, very helpful in thinking about their research methodology and ethics. Recognising the links between RDM and the

research process, they felt that producing a DMP should be part of the Project Approval. The students on the programme from other universities felt that the training was transferable to other university settings.

About Supporting infrastructure for RDM: Many published guidelines / materials exist to support RDM, however, much is targeted at information professionals, data managers etc, rather than research staff and students. To make this material more accessible, we created a customised Google search engine⁵ to a selection, rather than exhaustive list, of key resources. (The UKDA's models and materials are particularly useful as they are targeted at the researcher.) This provides the nucleus of one element of the infrastructure support required, however, clear policy, procedures and guidance are needed to help research staff / students and which identify roles and responsibilities. This will avoid them needing to recognise and / or interpret the sometimes different requirements (e.g. data retention periods that range from 3 to 10+ years across the seven Research Councils UK). A second RDM infrastructure element is tools and facilities, including software / hardware for data creation; secure data storage, access and retrieval; preservation and deletion. Whilst student participants were generally happy with the resources available to them to support RDM, some would have liked access to a simple 'data management system' to automate aspects such as file management, document versioning etc. Research staff requirements for tools and facilities are likely to be much more demanding, especially for collaborative projects. The third RDM infrastructure element is staffing and training; staff to support RDM in HEIs and to deliver training / education about RDM. HEI infrastructure to support RDM (i.e. policy and procedures, human and technical) may require further development to address the demands from external drivers and to meet the needs of its research staff and students.

Impact of the project

The pilot programme has already been valuable for participants and has had an impact. It has increased their awareness of RDM and how some RDM issues are closely related to issues concerning research methods and ethics; they feel more confident about planning and managing their research data. Some participants have completed their DMP and have used it to complete their project approval and ethical approval submissions.

In 2011/12 the training will be embedded within The Graduate School training programme for new and existing PGR students / supervisors at Northumbria and if possible offered to academic staff and to other students / staff in the region. It will also be extended to selected Masters / Doctoral programmes by piloting elements in research methods modules in the School of Computing, Engineering and Information Sciences.

The wider research community has also benefitted. Health information professionals, researchers and others in the region and beyond have learned about different approaches to RDM and digital presentation in the health sector from a range of practitioners (the Data4Life event - session 4 - on 26 May 2011). Northumbria researchers and the Advisory Panel have raised awareness and an increased understanding of the importance and relevance of RDM to their activities and their role in actively supporting effective RDM. New networks have been formed with the DCC, DPC and UKDA.

What we concluded

RDM should be seen as an integral part of 'good research', underpinning research excellence wherever the research takes place. The *DATUM for Health* project demonstrated that RDM training is valuable for PGR students, should be embedded in their training provision and also made available for research staff. The DMP should be viewed as a central component of the research process feeding into other essential components, viz. the research proposal, ethics, methodology and, particularly in the case of PGR students, the training plan. It should be a mandatory part of the doctoral study approval process. The DMP used on the project was customised for the target audience, and developed from that of the DCC. A one-size-fits-all DMP is probably not achievable or desirable. But training PGR

students to use a DMP, of whatever design, will develop good RDM principles and practices and enable them to easily deal with DMP templates they will encounter in their future research careers. Structuring a DMP into the students' PhD journey, relating it to milestones such as project approval and ethical approval, and connecting it to their personal development portfolio, would be of great benefit, and would enable good RDM throughout the PhD study.

Discipline-specific training (e.g. health-specific or similar disciplines grouped together) enables discussions to be more meaningful and useful, particularly on ethical issues which have a greater focus in health-related disciplines. Methodology-specific training is less critical since many researchers use mixed methods. Resourcing discipline-specific training is, however, costly and probably not sustainable. The DATUM team suspects that 80% of the training materials can be generic. A pragmatic and sustainable way of delivering the disciplinary focus and contextualisation is to 'tailor' generic materials through (a) discussion about research philosophy / epistemology; (b) covering specific requirements of qualitative or quantitative data; and (c) incorporating discipline-specific examples, case studies, exercises and references.

The long term storage of research data presents two infrastructural issues: (i) appraisal to decide what data is worth keeping and what can be destroyed, and (ii) where such data can best be stored. It is not appropriate to keep all research data for ever, nor is it necessarily appropriate to keep all data for 10 years after the end of a project (as recommended by some RCUK). Though data storage is cheap, data retrieval / (re)discovery is not and preservation of electronic data is expensive. Some data sets are small or topical or methodologically inappropriate for reuse, or are effectively published as the results. Additionally, without appraisal, repositories (local or national) would be totally swamped with data. An important component of any RDM training is therefore appraisal skills.

The location for data storage is also problematic. Data from PhD and academics' research could be deposited in University repositories. However, currently most of these repositories only store theses and outputs. New procedures and increased resources would be needed for these repositories to take on this new task. National repositories will be stretched to take on the task of receiving all the data from RCUK funded projects. This is also only a portion of all the research conducted in the UK, and does not include research funded by others, PhD research or scholarly activity. National repositories could not take on the task of storing all this data, particularly if proper appraisal is not undertaken. Enhanced publications could be more utilised to publish data that has a wider value but is not of sufficient value to be stored in a national repository.

What we recommended

A series of recommendations and actions emerged from the project for different stakeholders.

General recommendations to HEIs:

1. HEIs should adopt / adapt ideas from the RDMTrain projects: (i) using the models / approaches to implementation; (ii) using the training materials
2. HEIs should view the Data Management Plan (DMP) as being a central component of the research process feeding into other essential components, viz. the research proposal, ethics, methodology and, particularly in the case of PGR students, the training plan. These components should be supported by enabling infrastructure, both technical and non-technical.
3. HEIs should make RDM an integral part of their PGR training / education programme, not a separate programme (as was necessary in the DATUM for Health project pilot)
4. Learning about / exposure to the concept of RDM and the benefits of completing a Data Management Plan (DMP) should be mandatory for PGR students, e.g. as a component of a mandatory induction session

5. Other RDM specific training sessions, to provide the necessary knowledge and practical skills to implement RDM, should be offered
6. HEIs should ensure that RDM is noted, as applicable, in other PGR training sessions (e.g. ethics, DPA / FoI, copyright) and that reference to RDM specific training sessions is made
7. HEIs should recommend that a Data Management Plan (DMP) is an explicit requirement of the approval of a PGR student's study programme / proposal and aligned with the ethics approval system
8. HEIs should align the DMP with a PGR student's professional development portfolio, e.g. by making an explicit link to it being part of the Vitae RDF
9. HEIs should include an RDM component in a mandatory session for 2nd / 3rd year PGR students to train 'legacy' students (this would be a one-off activity)
10. HEIs should provide research staff / supervisors with an RDM training opportunity; this could be achieved by allowing staff to participate in the RDM sessions within their PGR training / education programme
11. HEIs should consider establishing a working group to develop a strategy and action plan for RDM in the university. This should include the review of existing infrastructure to support RDM (i.e. policy, procedures and guidance; tools and facilities; roles and responsibilities, staffing, training) in the light of external drivers. Such a group should include information / data management tutors; central research support staff, institutional repository and records management / data protection / freedom of information staff; academic / research staff.
12. RDM training should be extended across an HEI via, for example: an awareness raising / training session(s) for PGR supervisors; an awareness raising / training session(s) for academics / research staff; embedding RDM into taught PG research methods modules, including MRes and Professional Doctorate; embedding RDM into undergraduate research methods modules; an awareness raising / training session(s) for support staff, e.g. research administrators, IT staff. These can be achieved by using and / or tailoring the DATUM training programme materials and materials from the other RDMTrain projects.

Recommendations for the wider community

To Research Funders

1. To agree on a standard DMP template for all research council funded projects
2. To establish clear, justified guidance on appraisal and a retention schedule
3. To explore other mechanisms for making research data widely available, e.g. promoting enhanced publications: for small projects this would be an easier, more practical method than repositories; this method might fit better into the REF system

To Data Management Services

1. The UKDA is an exemplar of good practice and guidance
2. Produce clear guidance and training materials on appraisal. It is not practical to keep all research data (though storage might be cheap, (re)discovery and preservation is expensive). It is clear that some data can be destroyed at the 'end' of the project (e.g. small data sets, or very topical data); other data is of such significance that it should be placed in a repository (e.g. large scale studies, work of leading researchers, topics of historical significance). Appraisal guidance is urgently needed for the data that falls between these two extremes.

To JISC

1. JISC should recommend to Vitae that RDM and a DMP be an explicit element of their Researcher Development Framework
2. JISC should consider supporting the training of the 'legacy' of PGR supervisors / researchers rather than individual HEIs e.g. by funding / organising regional training events during 2012; collaborating with Vitae in relation to PGR training

3. JISC should consider funding a gateway to all the RDM resources that exist, either as literature or as Web resources, for example by developing further the DATUM RDM Google search engine. The starting point would be the resources from the MRD programme projects.

¹ This is the case for the BBSRC, ESRC, MRC, NERC and the Wellcome Trust.

See: BBSRC (2010). BBSRC Data Sharing Policy. <http://www.bbsrc.ac.uk/web/FILES/Policies/data-sharing-policy.pdf>; ESRC (2010).

ESRC Research Data Policy. http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf; MRC (n.d.)

MRC Policy on Data Sharing and Preservation.

www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/Policy/index.htm; NERC (2011).

NERC Data Policy. www.nerc.ac.uk/research/sites/data/policy.asp

Wellcome Trust (2010) Policy on Data Management and Sharing. <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>

² JISC Managing Research Data Programme <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>
Research Data Management Training Materials (RDMTrain)

<http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain.aspx>

³ See full project report on the project website for details of how the project was conducted

<http://www.northumbria.ac.uk/datum>

⁴ See training materials on the project website for full details of the programme content and materials which are available for use <http://www.northumbria.ac.uk/datum>

⁵ Research Data Management Training custom search engine <http://goo.gl/aqVNQ>