



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

## **JISC Final Report**

Julie McLeod, Northumbria University

August 2011

Funded by JISC, under the Managing Research Data (JISCMRD) Programme: Promoting discipline-focused research data management skills

**This work by Authors:**

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

**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

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

Dr 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>

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

© Copyright holder: Northumbria University

**METADATA**

<b>Creation</b>	
Creator names	Julie McLeod
Date of creation	30 June 2011
Status	Draft
<b>Amendments</b>	
Version creator names	Julie McLeod
Version number	2.2
Version date	2 Aug 2011
Status	Final
<b>Type of document</b>	Report
<b>Subject keywords</b>	Research data management; training; health discipline; DATUM project

## Table of Contents

<b>1</b>	<b>ACKNOWLEDGEMENTS .....</b>	<b>4</b>
<b>2</b>	<b>PROJECT SUMMARY .....</b>	<b>4</b>
<b>3</b>	<b>MAIN BODY OF REPORT .....</b>	<b>5</b>
3.1	PROJECT OUTPUTS AND OUTCOMES .....	5
3.2	HOW DID YOU GO ABOUT ACHIEVING YOUR OUTPUTS / OUTCOMES? .....	6
3.3	WHAT DID YOU LEARN? .....	8
3.4	IMMEDIATE IMPACT .....	9
3.5	FUTURE IMPACT .....	9
<b>4</b>	<b>CONCLUSIONS .....</b>	<b>10</b>
<b>5</b>	<b>RECOMMENDATIONS .....</b>	<b>11</b>
5.1	GENERAL RECOMMENDATIONS TO HEIS .....	12
5.2	RECOMMENDATIONS FOR THE WIDER COMMUNITY .....	12
5.3	RECOMMENDATIONS FOR JISC .....	13
<b>6</b>	<b>IMPLICATIONS FOR THE FUTURE .....</b>	<b>13</b>
<b>7</b>	<b>REFERENCES .....</b>	<b>13</b>
<b>8</b>	<b>APPENDICES .....</b>	<b>15</b>
8.1	DATUM FOR HEALTH PILOT PROGRAMME - EVALUATION SUMMARY .....	15
8.2	FEEDBACK FORM GIVEN TO PARTICIPANTS AFTER EACH SESSION .....	16
8.3	TELEPHONE INTERVIEW QUESTIONS WITH VOLUNTEER PARTICIPANTS AFTER THE END OF THE PROGRAMME .....	17

## 1 Acknowledgements

DATUM for Health was funded through the JISC Managing Research Data programme <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx> as part of the Research Data Management Training materials (RDMTrain) strand <http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain.aspx>

The project team was a collaboration between two academic departments at Northumbria University (health studies and information management) and the University's Graduate School, with input from the Digital Curation Centre (DCC <http://www.dcc.ac.uk/>), Digital Preservation Coalition (DPC <http://www.dpconline.org/>) and an internal Advisory Panel.

Northumbria University project team members were:

- Professor Julie McLeod, School of Computing, Engineering & Information Sciences (Project Lead)
- Professor Charlotte Clarke, School of Health, Community and Education Studies
- Professor John Dean, The Graduate School
- Sue Childs, School of Computing, Engineering & Information Sciences
- Elizabeth Lomas, School of Computing, Engineering & Information Sciences

The team would particularly like to acknowledge the contributions of:

- Dr Kevin Ashley (Director, DCC) for supporting the collaboration and Ms Joy Davidson and Ms Sarah Jones, DCC for delivering one of the sessions in the training programme
- Dr William Kilbride (Executive Director, DPC) for supporting the collaboration and presenting at the final session of the training programme (Data4Life event); Ms Carol Jackson, DPC for her work helping to organise and ensure the smooth running of the Data4Life event
- Ms Laura Smailes, 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
- Presenters at the Data4Life event (<http://www.dpconline.org/events/details/28-data4life?xref=27>)

## 2 Project Summary

This collaborative project sought to promote the research data management (RDM) skills of postgraduate research (i.e. doctoral) students in the health studies discipline via a specially-developed training programme. It focussed on the management of qualitative, unstructured data. The training programme comprised four sessions, delivered over a period of weeks, and covered both generic and discipline-specific issues. Northumbria University delivered two sessions; the DCC delivered a tailored session on digital curation, and the DPC jointly hosted a roadshow session on digital preservation in health which was opened up to a wider audience. 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).

Participants were asked to evaluate the usefulness and effectiveness of the training programme. They rated it very highly, finding it enjoyable, useful and professionally run. This was true for each session and for the programme as a whole. The key outputs of the project are:

- a customised Google search engine providing access to a selected set of ~200 RDM Web based resources (RDM Training <http://goo.gl/aqVnQ>)
- a model for RDM skills training, based on andragogic principles
- a training programme for RDM in the health studies discipline (including materials and tutor notes) covering qualitative, unstructured data

- a project report containing: (i) details of how this training is being embedded in Northumbria University, and (ii) recommendations to HEIs, JISC, research funders and data management services
- a schedule for embedding RDM in the PGR training programme at Northumbria University for the 2011-12 academic year

The training model, programme and materials, and the project report will be made freely available on the Web (<http://www.northumbria.ac.uk/datum>) and via JORUM.

### 3 Main Body of Report

#### 3.1 Project Outputs and Outcomes

Output / Outcome Type (e.g. report, publication, software, knowledge built)	Brief Description and URLs (where applicable)
<b>Outputs:</b>	
Literature review	Over 200 items reviewed <a href="http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/litrev/">http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/litrev/</a>
A customised Google search engine	Provides access to the Web-based resources identified from the literature review for use in teaching, learning and research completed: <i>Research Data Management Training</i> <a href="http://goo.gl/aqVnQ">http://goo.gl/aqVnQ</a>
Training programme model	Based on the andragogic model of learning, incorporating presentations, group activities and discussion, directed learning and examples tailored to the discipline context
Training programme	A RDM training programme for the health studies discipline covering qualitative, unstructured data. 4 sessions with aims, learning outcomes & content: 1 - Introduction to Research Data Management 2 – Digital Curation (based on DCC’s 101 Lite) 3 - Problems and Practical Strategies and Solutions 4 – Data4Life – Northumbria University/Digital Preservation Coalition Roadshow Programme materials including tutor notes. Freely available via JORUM and the project website <a href="http://www.northumbria.ac.uk/datum">http://www.northumbria.ac.uk/datum</a>
Final report	Summarising the project’s methods, pilot programme evaluation, findings and conclusions, with recommendations on how to embed such research training into other (academic and staff) programmes; requirements for further research data management training and review of infrastructure to support research data management in an institution
Presentation(s)	<ul style="list-style-type: none"> <li>• ‘DATUM for Health- Promoting research data management skills in health studies’, Northumbria University Research Conference, 5-6 May 2011</li> <li>• ‘Developing Research Data Management Skills of Doctoral Researchers’. Workshop proposal, with other JISC funded RDMTrain projects, at Vitae Researcher Development International Conference 2011: Realising the potential of researchers, 5-6 September 2011 <a href="http://www.vitae.ac.uk/policy-practice/393351/Vitae-researcher-development-international-conference-2011-realising-the-potential-of-researchers.html">http://www.vitae.ac.uk/policy-practice/393351/Vitae-researcher-development-international-conference-2011-realising-the-potential-of-researchers.html</a></li> </ul>
<b>Outcomes</b>	
Embedding of training at Northumbria University	The SCEIS members of the Project Team will be offering RDM sessions (using the DATUM for Health model and materials) as a standard part of the PGR training programme at Northumbria University in the 2011-12 academic year
Knowledge built in the project team	More detailed knowledge of: RDM requirements, guidance/models and services (e.g. UKDA); RDM issues in health research; RDM approaches/solutions; health research issues and projects in the university/other regional universities/elsewhere
Networks/ collaboration	Established new, strong networks/collaboration: internally with The School of Health, Community & Education Studies and The Graduate School; externally with DCC, DPC and UKDA

### **3.2 How did you go about achieving your outputs / outcomes?**

The aims of the project were to promote research data management skills in Higher Education Institutions (HEIs) by:

- designing and piloting a training programme on RDM for postgraduate research (PGR) (i.e. doctoral) students in the health studies discipline, as an integral part of a doctoral training programme. The programme covered both generic and discipline-specific issues, focussing on the management of qualitative, unstructured data
- evaluating the usefulness and effectiveness of the training programme with the PGR student participants and other research stakeholders
- providing other HEIs with a model for research data management skills training
- making recommendations on: how to embed such research training into other (academic and staff) programmes; requirements for further research data management training and a review of infrastructure to support research data management in an institution.

The project was conducted in seven phases:

1. A targeted literature review  
this was conducted to identify and review literature to (a) collate best practice guidelines and RDM requirements (generic, for qualitative, unstructured data, and for the health studies discipline); (b) identify training/learning delivery models for the range of stakeholders; and (c) identify learning materials for tailoring to the target audience. Since many guidelines and short training courses on RDM already exist, the training drew on them, tailored them for use and only developed further guidance/material to fill gaps. The outputs were a collection of existing Web-based resources (relevant to the project's focus), which were made searchable via a customised Google search engine (<http://goo.gl/aqVNQ>), and a summary of the literature review<sup>2</sup>.
2. Design of the training programme  
This comprised both the model and material type(s). A questionnaire survey was conducted with PGR students in health (6 respondents), looking at what data they collected, how they managed their data at all stages in the lifecycle, how well they thought they managed their data, if they would like to have training, and the approach and materials such training should comprise. A focus group was held with three other research stakeholders in the School of Health, Community & Education Studies (SHCES) and one from another school. The focus group explored their experiences of RDM - what worked well, what could have been better - and what kind of training was required for data management. This enabled the preferences and already known needs of the target audience to be identified as well as the potential unknown needs (as expressed by other experienced stakeholders). The output was a model and draft outline for a training programme for Phase 3.
3. Development of the training programme  
The team, including the experts from the DCC/DPC, drew upon their knowledge and skills in research data/information management, their expertise and experience in teaching and learning and relevant existing materials, appropriately tailored, to develop the programme. The andragogic model of teaching underpinned the programme's design<sup>1</sup>. If researchers are to execute their responsibilities effectively, and be equipped to continue to acquire more in-depth skills as appropriate, then an andragogic approach to training programme design is required. This approach places the learner (in this case the researcher) at the centre rather than the tutor. Adult learners want to: understand why they are undertaking a learning activity and therefore become internally motivated; base their learning in their previous experiences; be involved in planning and evaluating their learning activities; have learning activities that are immediately relevant to their work and are problem, not content, oriented. Activities included: group sessions with individual students sharing their prior experience and knowledge and what they want from the programme (addressing need to know, experience and planning in the andragogy model); directed learning, e.g. completing their own data management plan and discussing this in the group setting (addressing problem-centred activities and relevance to their work in the andragogy model); feedback forms after each session and telephone interviews with volunteers to evaluate the programme (addressing evaluation in the andragogy model).

The programme covered all of the stages of the data lifecycle and comprised four sessions, building upon each other (see 3.1 above). It was also designed to dovetail with existing research

skills/research management provision in the university, in particular relating to PGR milestone preparation (Initial Project Approval and Mid-Point Progression), records management, DPA/Fol/copyright, ethics and governance. Good andragogic / pedagogic practice was used to develop a format, particularly for sessions 1 and 3, that comprised a mix of lecturing, real-life examples, worked exercises, group work, class discussion and directed learning<sup>1</sup>. The DCC's digital curation session was also used to 'train the trainers' (i.e. build expert capacity), so that in future this session will be delivered in-house. Similarly, the DPC roadshow style event supported capacity building so that the digital preservation element will be delivered in-house in the future. This roadshow was innovative for the DPC being a subject specific event, and has extended the scope of such DPC events for the future.

The varied nature of research, even within a single discipline such as health studies, means that learning in one place at one time is not always convenient or possible. PGR students cannot always attend training sessions (due to the nature and constraints of their research) therefore sessions were audio recorded for use at different times and in different places. Materials were stored on the virtual learning environment (VLE) and were in different formats to support different cognitive styles viz. PowerPoint presentations with audio, notes of the discussion, templates, a customised search engine, videos from the Incremental project, DCC/DPC/UKDA leaflets/documents.

#### 4. Piloting and evaluation of the training

The programme was advertised to PGR students in health studies at Northumbria University and other regional Universities who are members of FUSE (a UK Centre for Translational Research in Public Health <http://www.fuse.ac.uk/>). 25 people attended some component of the programme. 19 students were from the School of Health, Community & Education Studies (SHCES) at Northumbria University, with one from another school and two from other regional universities; three researchers also took part. Because of the legacy of students and staff who have not received prior skills training in this area, participants were at all stages of their doctoral study / research careers. One student was visiting from overseas.

Participants were asked to evaluate each training session, using The Graduate School's standard PGR training session feedback form. This included questions about what were the most useful/least useful aspects of the session; what did they want to know more about; the pace, content and presentation of the session; and suggestions for any improvements that could be made or any further training they would like. In addition, for sessions 1 to 3 the project team and the DCC self-evaluated the sessions, reflecting on what went well and what could have been better. The evaluation responses were used to inform the delivery of subsequent sessions. At the end of the programme, we asked for volunteers from the students to be interviewed by phone. These interviews (6 in total) explored the programme as a whole, and asked questions about the appropriateness of the format of the pilot, how future delivery of the programme could be structured, and what additional resources were needed to be able to manage their data more effectively. The Advisory Panel, comprising a wide range of stakeholders, plus the focus group participants from Phase 2, were asked to give feedback on the draft recommendations for embedding and extending RDM training at the University. This informed plans for transferability and wider recommendations for future action.

#### 5. Refinement of the training programme and publication

Based on the project team tutors' experience of delivering the programme and its evaluation, the materials were refined for publication on the Web and deposit into JORUM. The final materials comprise: the training model; the training programme and rationale for its design; PowerPoint slides with notes for tutors; materials for learning activities with notes to support the activity and discussion; support materials including a list of those used and freely available elsewhere.

#### 6. Embedding the training programme into the university's wider training programme for postgraduate researchers

The project team presented the Advisory Panel with a set of ideas based on the evaluation (from participants and project team members). These were refined and a set of actions was agreed to (a) embed the training within The Graduate School training programme for new and existing PGR students/supervisors from 2011/12 and (b) extend the training at Northumbria to selected Masters/Doctoral programmes. The latter will be achieved by piloting elements within other

research methods modules on Masters distance learning programmes and a new Professional Doctorate programme in the School of Computing, Engineering and Information Sciences.

#### 7. Making recommendations for sustainable RDM training and associated infrastructure requirements

Draft recommendations were discussed at the final Advisory Panel meeting and refined into an action plan for presentation to Northumbria University.

### 3.3 What did you learn?

- RDM guidance

Many guidelines and materials exist to support RDM. Whilst this is very useful for information professionals, data managers etc, it can be daunting to data creators and users (in this case PGR students). As a way of making such material more accessible, we created a customised Google search engine <http://goo.gl/aqVnQ> and invited other RDMTrain projects to contribute to this. This search engine is not exhaustive, but provides a nucleus that could be developed further by JISC or other interested parties. From the literature review the UKDA's models and materials, available through their 'Create and manage data' web pages (<http://www.data-archive.ac.uk/create-manage>), were found to be particularly useful as they were targeted at the researcher rather than the information professional. The Project Team visited the UKDA and had a very productive meeting exchanging ideas. Information gained there about the UKDA's training materials led to their use in Session 3 of our programme on practical issues and solutions.

- RDM pilot training programme

25 people (22 PGR students and 3 academic researchers) attended at least one session of the training programme; 11 attended three or four sessions. In their evaluation participants rated the pilot programme very highly, finding it enjoyable, useful and professionally run. This was true for each session and for the programme as a whole.

The flow and the logic of the programme were felt to be good. The programme was designed in a holistic way, each session building on from the previous one, with directed learning activities to consolidate learning/prepare for the next session. Interestingly, participants who missed some sessions did not feel that they were disadvantaged; the sessions could also 'stand alone', and attendance at even one of them would be beneficial. It was felt that three sessions would be sufficient in the future. This will be achieved by incorporating elements of digital preservation covered in the final session into the second session on the digital curation lifecycle. The pattern of the first three sessions, which comprised lectures, worked exercises, group work and discussion and directed learning, was liked. Students also enjoyed the full day event (final session 4) which gave a wider perspective of RDM from a range of practitioners. Face to face sessions were preferred. Only 11 students joined the DATUM for Health organisational site on the VLE; this mirrors the experience of The Graduate School, with PGR students not being big users of the VLE.

RDM specific sessions were thought to be necessary, though RDM could also be covered in other courses in the PGR training programme, e.g. induction. Participants felt an important time for the sessions should be early in the PhD journey, e.g. before project approval, though RDM training would be useful at other stages too. Session length of 2.5 hours was thought to be appropriate, and sessions should be at weekly to monthly intervals to allow time for carrying out the directed learning tasks. A customised data management plan (DMP) was developed for the programme based on that of the DCC. The students used it in a worked exercise and group discussion. This was followed up with a directed learning activity to complete the DMP for their own research project. Participants found this DMP activity very helpful. They felt that producing a DMP should be part of the PhD study's project approval process. The potential for 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, makes the DMP a living document following the data lifecycle. Some students on the programme used the DMP in their project approval and ethical approval submissions.

Participants preferred discipline-specific training (e.g. health-specific or similar disciplines grouped together (e.g. health, social sciences and humanities)). Most were not concerned about methodology-specific training (i.e. covering qualitative or quantitative data only). This could be because many health researchers use mixed methods, and the approach to quantitative data is strongly influenced

by the discipline of the researcher. The focus on the health discipline and qualitative data clearly demonstrated that RDM is inextricably related to research methodology and to ethics. These topics are covered in other sessions on the University's PGR training programme. It is clearly important to relate these sessions to each other so the student can see the connections between these topics, but also how there are different perspectives on RDM.

Students found Session 3 on practical solutions, which included file naming conventions and 'file plans', very useful. Participants were generally happy with the resources available to them to support RDM. However, a few students would have liked access to a simple 'data management system' to automate aspects such as file management, document versioning etc.

The students on the programme from other universities felt that the training was transferable to other university settings.

### **3.4 Immediate Impact**

PGR students said the programme had increased their awareness of RDM and how some RDM issues are closely related to issues concerning research methods and ethics. Some said they felt more confident about planning and managing their research data. There is evidence from the SCHES that some have completed their DMP and have used this to complete their ethics review documents (submitted to the school's Ethics Committee for approval) and, in one case, to prepare for their formal Mid-Point Progression meeting. One of the external students used the DMP at their 'project approval' stage. Feedback from SCHES staff is that students have spoken about the programme's value and usefulness, for example with their supervisors and at a recent research 'away day'. The Graduate School are very enthusiastic about embedding the programme into their well-established PGR training programme, available to all PGR students. They are also keen to extend the opportunity to attend to academic staff and to other students/staff in the region. Actions have been agreed to ensure this happens in the next academic year (2011/12).

The wider research community benefits have been:

- For Northumbria researchers, a raised awareness of the importance of RDM and of JISC's MRD programme (presentation at Northumbria University's Research Conference, May 2011)
- For health information professionals/researchers/others in the region and beyond, learning about different approaches to RDM and digital presentation in the health sector from a range of practitioners (the Data4Life event 26 May 2011)
- For Northumbria University's project team members in SCEIS a new network with UKDA staff and potential future collaboration
- For Advisory Panel members working in the University repository and research support, an increased understanding of the importance and relevance of RDM to their activities and their role in actively supporting effective RDM.

Evidence of this impact has come from the formal evaluation of the programme (feedback forms/interviews) and ad hoc/informal comments at the training sessions, to colleagues and fed back to the project team.

### **3.5 Future Impact**

RDM training will become a standard part of the PGR training programme at Northumbria University in academic year 2011-12. This training will comprise a component in the compulsory induction sessions for new PGR students, plus two other RDM-specific sessions. The latter (a recommended session on data management planning and the data lifecycle and an optional session on practical implementation of RDM) will be made available to PGR students at *any* stage in their studies, as well as to researchers and academics. The PGR programme will also be available to students at other Universities in the region, who will be given the chance to enrol on the RDM sessions.

A recommendation is being made to The Graduate School that a DMP should be made a required component of the PGR project approval process. If accepted, a regulation change would occur in academic year 2011-12, with implementation in 2012-13.

RDM training will be extended to selected postgraduate programmes in SCEIS in academic year 2011-12. This will comprise including RDM/DMPs in research methods modules on Masters distance learning programmes and the new Professional Doctorate in Information Science. These pilots will be used to inform future plans for extending RDM training across the University

Impact of the above activities will be tracked via formal evaluation of the programme sessions (feedback forms); evidence of RDM plans/planning influencing ethics review forms and project approvals of PGR students; increased/better RDM via DMP implementation, monitored through the PGR annual progression monitoring process.

Northumbria University will extend training in the region by advertising the PGR training sessions to the other four HEIs in the North East (Durham, Newcastle, Sunderland and Teesside). In addition the following will be explored:

- offering a regional training day on RDM via the Vitae Regional Hub
- including RDM in the activities of the North East Collaboration Group for Researcher Development
- providing a training day, on request and with payment, to other universities in the country.

Impact will be tracked by the use of session feedback forms.

A session on research data and digital preservation will be included in the Northumbria University Research Conference 2012 which is open to all staff (academics, researchers, support). The aim of this session is to raise awareness and increase engagement in RDM. Impact will be tracked by the use of session feedback forms.

The Project Team are writing internal reports and recommendations targeted to Northumbria University committees, e.g. The Graduate School, the Research Committee, and the Research Business & Innovation department. These reports will be tabled in the 2011-12 academic year. Impact will be tracked via monitoring of changes in University procedures and systems.

The project's teaching materials will be publicly available, with the aim that others will make use of these. Impact will be tracked via Web usage statistics and ad hoc feedback.

## 4 Conclusions

The evaluation of the pilot programme showed that PGR students found the style, content and delivery to be very useful and valuable, irrespective of the stage of their study. However, they felt it would be better to have this training at an early stage of their PhD. They may not all have had knowledge of or an appetite for RDM at the start of the programme but all quickly recognised its importance as part of doing good research. Most preferred discipline-specific training e.g. health-specific or similar disciplines grouped together, for instance health, social sciences and humanities. This enabled the discussions to be more meaningful and useful, particularly on ethical issues as health-related disciplines have more of a focus on and knowledge of ethics. Most were not concerned about methodology-specific training, i.e. focusing only on qualitative or only on quantitative data. They were happy to cover both types of data as many qualitative 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<sup>1</sup>. 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. [Examples of how this was done in the DATUM project training materials include: discussing ethical issues in health research and the implications of being able to share and/or re-use data; the validity of using/re-using data qualitative collected in a particular context, for a particular purpose and by/from particular individuals (an epistemological issue); the use of a health information research project for the sample DMP exercise.]

---

<sup>1</sup> This is an estimate only based on the Pareto principle or 80:20 rule  
[http://en.wikipedia.org/wiki/Pareto\\_principle](http://en.wikipedia.org/wiki/Pareto_principle)

Participants found the data management plan (DMP) exercise and associated directed learning particularly helpful – a framework for ensuring they managed their research data. They felt that producing a DMP 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.

Good RDM processes are required wherever 'research' takes place, from undergraduate to postgraduate and doctoral students, researchers, academics and research administrators. Training therefore needs to be extended to cover RDM in research methods modules at all levels (i.e. undergraduate, taught postgraduate, MRES, Professional Doctorates), as appropriate and relevant to the level of study. In addition, awareness raising sessions on RDM need to be provided in staff development programmes for both academics and support staff.

The keen interest and high level of engagement with the project by the programme participants, the Advisory Panel and other stakeholders demonstrate that JISC's aim to promote and support good research data management seems to be 'knocking at open doors'. In a few years, it is possible that at least some universities will be providing embedded RDM training for new students (from undergraduate level upwards) and staff. However, there is a large legacy of current students, researchers and academics that have yet to receive any RDM training. Such legacy training is likely to take many years to achieve and be extremely costly.

The 'elephant in the room' is the long term storage of research data. There are two aspects to this issue, (i) appraisal to decide what data is worth keeping and what can be destroyed, and (ii) where such data can best be stored. Appraisal of research data is an issue that has not been sufficiently addressed. 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 RCUK). Though data storage is cheap, data retrieval / (re)discovery is expensive, and the more data there is to sift the more expensive it becomes; preservation of electronic data is also expensive. Some data sets are small or topical or methodologically inappropriate for reuse (particularly true for some qualitative research methodologies), or are effectively published as the results. In such cases, data would only need to be kept until the end of the project and publication of the findings. Additionally, without appraisal, repositories (whether 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 Research Council 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.

All these conclusions are general and relevant to the wider community.

## 5 Recommendations

Good RDM processes are required wherever 'research' takes place, from undergraduate to postgraduate and doctoral students, researchers, academics and research administrators.

## **5.1 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 project's 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. 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
6. 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
7. 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
8. HEIs should include an RDM component in a mandatory session for 2<sup>nd</sup>/3<sup>rd</sup> year PGR students to train 'legacy' students
9. 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
10. 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.
11. RDM training should be extended across an HEI via, for example: an awareness/training session(s) for PGR supervisors; an awareness/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/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.

## **5.2 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.

### 5.3 Recommendations for 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.

## 6 Implications for the future

The project has already resulted in change at Northumbria University. The SCEIS members of the Project Team will be offering RDM sessions as a standard part of the PGR training programme at Northumbria University in academic year 2011-12. Additionally, they will be piloting adding RDM coverage into some taught postgraduate research modules in SCEIS in 2011-12. The aim to extend RDM awareness raising and training across the University will be met by recommending that a working group be set up to prepare a medium-longer term strategy for RDM across the University. This working group should review existing infrastructure to support RDM (i.e. policy, procedures and guidance; tools and facilities; roles and responsibilities; staffing) as well as looking at training for all staff and students, and how this can be delivered in a sustainable fashion.

Knowledge about the DATUM for Health project is spreading across the region, and there are plans to hold a regional RDM awareness raising / training event in 2011-12.

As well as making the teaching materials and project report publicly available, the SCEIS Project Team members will be writing a journal article to share their learning from the project to a wider audience.

The major implication for the wider community is the sustainability of implementing RDM training and education. Sustainability issues include:

- availability of RDM subject expertise for delivering the teaching
- workload for tutors
- the huge number of research/support staff needing awareness raising/training sessions
- resourcing training materials development, staff to deliver training, staff time to attend training
- the long time lag before RDM could become embedded into taught PG and UG research methods modules.

Long term project contacts:

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

[julie.mcleod@northumbria.ac.uk](mailto:julie.mcleod@northumbria.ac.uk)

Professor John Dean, Director of The Graduate School [john.dean@northumbria.ac.uk](mailto:john.dean@northumbria.ac.uk)

Programme and training materials will be made available via the project web site

<http://www.northumbria.ac.uk/datum> and via JORUM. The project website will be maintained as part of Northumbria University's website. The project website has been offered to the British Library UK Web Archive<sup>3</sup>.

There are no resources to extend the Research Data Management search engine <http://goo.gl/aqVnQ> after the end of the project, though SCEIS project team members will continue to add in further items on an ad hoc basis. The administration rights to the search engine could be transferred to another group/organisation to maintain its currency.

## 7 References

1. Knowles MS, Holton EF & Swanson RA. (2005). *The adult learner: the definitive classic in adult education and human resource development*. 6th ed. Elsevier; Paraskevas A & Wickens E. Andragogy and the Socratic method: the adult learner perspective. *J. Hospitality, Leisure, Sport &*

- Tourism Education*, V2 (2), 2003. <http://www.heacademy.ac.uk/johlste>; Race, P (2010) *Making learning happen*: 2nd edition London: Sage Publications.
2. The summary of the literature review undertaken for the project available on the project website <http://www.northumbria.ac.uk/datum>
  3. DATUM for Health project website: <http://www.northumbria.ac.uk/datum> (Offered to the British Library UK Web Archive <http://www.webarchive.org.uk/ukwa/>)

## 8 Appendices

### 8.1 DATUM for Health Pilot Programme - Evaluation Summary

#### Evaluation approach

- Feedback forms from participants after each session
- Self evaluation by presenters after each session: what went well; what could have been better
- Phone interviews with volunteer participants after end of programme (6 people interviewed)

#### Student numbers and characteristics

The pilot programme comprised four sessions, building upon each other. 25 students took some component of the programme:

- Session 1 = 20 students
- Session 2 = 13 students
- Session 3 = 13 students
- Session 4 = 11 students

Total number of individual sessions attended:

- All 4 sessions = 5 students
- 3 sessions only = 6 students
- 2 sessions only = 5 students
- 1 session = 9 students

There was no pattern to the sessions attended / not attended, e.g. some students only attended the first session, others didn't attend this session but attended later ones, etc. For some students, who sent apologies, there were other commitments on the days of a given session. For other students it could be that some sessions seemed more relevant to them than others.

Student characteristics:

- Full time PGR student = 16
- PT PGR student = 6
- 1<sup>st</sup> year PGR student = 13
- 2<sup>nd</sup> year PGR student = 6
- 3<sup>rd</sup> year PGR student = 3
  - All from the School of Health, Community and Education Studies, except 1 student from another School and 2 students from other universities
- Academic = 3 (from the School of Health, Community and Education Studies)
- Users of DATUM organisation site on VLE = 11 (9 FT PGR students, 2 PT) Note: external students were emailed all the materials that were placed on the VLE organisation site

#### Evaluation findings

- Participants rated the course very highly: enjoyable, useful, and professionally run. This was true for all the sessions, and for the course as a whole.
- Constructively critical comments were minor, and will be used to tweak the course materials and inform the plan for embedding the training at Northumbria University.
- The flow and the logic of the programme were felt to be good: it is probable that three sessions will be sufficient in the future (the first three sessions provided in the pilot).
- The sessions should be timed early in the PhD study, before the project approval stage, though they would also be suitable for students at later stages in their PhD. A session length of 2.5 hours is appropriate. Sessions should be at weekly to monthly intervals to allow time for the directed learning tasks to be completed.
- The mix of presenters (academics, external practitioners, a PhD student) was liked; the ability to learn from people with expertise and experience in the topic, who answered queries and engaged in discussion was felt to be beneficial. The case studies presented in the roadshow session were found to be very interesting; though the format of a whole day conference was not thought appropriate by a number of the participants.

- The pattern of the first three sessions, which comprised lectures, worked exercises, group discussion and directed learning, was liked. The handouts and materials supplied were found to be very useful.
- Research data management (RDM) specific sessions were thought to be necessary, though RDM could also be covered in other courses in the PGR training programme, e.g. induction.
- Participants preferred discipline-specific training (e.g. health-specific) or similar disciplines grouped together (e.g. health, social sciences and humanities). Most were not concerned about methodology-specific training (i.e. covering qualitative or quantitative data only).
- Face to face sessions seem to be preferred. The group discussions, where students (with different backgrounds and at different stages in their PhDs) exchanged experiences, were felt to be particularly beneficial. PGR students do not appear to be big users of the VLE: only 11 students joined the DATUM for Health organisational site.
- Some participants thought that RDM training should be mandatory for PGR students.
- Participants found the data management plan (DMP) exercise and associated directed learning very helpful. They felt that producing a DMP should be part of the project approval process.
- Participants were satisfied, on the whole, with the resources available to them for RDM. Access to resources can be more of a problem for PT students. Access to secure physical storage can be a problem for some participants. One participant wanted access to a simple 'data management system'.
- The external students felt the training was transferable to other University settings.

## 8.2 Feedback form given to participants after each session

PGR Skills Training Programme 2010/11				
<i>Course Title:</i>				
<i>Date:</i>				
<i>Facilitator(s):</i>				
1. What were the most useful aspects of the workshop?				
2. What were least useful?				
3. What do you want to know more about?				
4. Was the pace of the workshop: <i>(please circle)</i>				
Too Slow          Just Right          Too Fast				
5. Was the content of the workshop: <i>(please circle)</i>				
Excellent          Good          Acceptable          Poor				
6. Was the presentation of the workshop: <i>(please circle)</i>				
Excellent          Good          Acceptable          Poor				
7. Please suggest any improvements, not mentioned above, that could be made to the workshop or any further training that you would like.				

### **8.3 Telephone interview questions with volunteer participants after the end of the programme**

Looking at the programme as a whole:

1. Did it work as a coherent programme?
2. Was there anything that did NOT need to be included?
  - a specific session, particularly the DPC one
  - a specific topic
3. Were there any gaps? Anything you would have liked to be covered?
4. What did you think of the format?
  - Timing in PGR study, e.g. at start of study, at other points during study
  - Number of sessions
  - Length of sessions
  - Presenters of sessions, e.g. different types of presenter (academics, PhD student, external organisations)
  - Style of sessions: lecture, examples, questions and answers, worked exercises
  - Directed learning tasks
  - Handouts
5. Blackboard organisational site
  - Did you use the site?
  - If not, why not?
  - If yes, was it useful? Could it be improved?
6. Future delivery of the programme:
  - face to face sessions
  - DL resources as a supplement to face to face
  - DL provision only
  
  - stand alone
  - embedded within other PGR training courses, e.g. ethics, data protection, project management
  
  - discipline specific classes, e.g. health related PGR students
  - discipline mixed classes, e.g. any and all PGR students
  
  - research methodology specific classes, e.g. qualitative only, quantitative only
  - research methodology mixed classes, e.g. both qualitative and quantitative
7. Should research data management training be mandatory?
8. Would it be beneficial for students to produce a Data Management Plan as part of the IPA?
9. Are there any resources, additional to those already provided, that you need to manage your research data?
10. Any other comments?