DATUM in Action
Supporting researchers to plan and manage their research data

JISC Final Report

Julie McLeod, Northumbria University

April 2012

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

This work by Authors:
Julie McLeod, Northumbria University, School of Computing, Engineering & Information Sciences
Sue Childs, Northumbria University, School of Computing, Engineering & Information Sciences

Project Lead

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

Project Team

Sue Childs, School of Computing, Engineering & Information Sciences
Elizabeth Lomas, School of Computing, Engineering & Information Sciences
Professor Maia Angelova, School of Computing, Engineering & Information Sciences
Professor Glenda Cook, School of Health, Community and Education Studies
Professor Anthony Hildreth, School of Health, Community and Education Studies
Dr Jeremy Ellman, School of Computing, Engineering & Information Sciences.

Project websites

Northumbria University: http://www.northumbria.ac.uk/datum
JISC http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx
http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure.aspx

METADATA

Creation	
Creator names	Julie McLeod
Date of creation	30 Mar 2012
Status	Draft
Amendments	
Version creator names	Julie McLeod
Version number	2
Version date	13 Jul 2012
Status	Final
Type of document	Report
Subject keywords	Research data management; data management plans; guidance; infrastructure; DATUM project

Table of Contents

1 Acknowledgements

DATUM in Action was funded through the JISC Managing Research Data programme http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx as part of Strand B http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure.aspx

The project team was a collaboration between two academic departments at Northumbria University (health studies and information management) and an internal Advisory Panel.

Northumbria University project team members were:
- Professor Julie McLeod, School of Computing, Engineering & Information Sciences (Project Lead)
- Sue Childs, School of Computing, Engineering & Information Sciences
- Elizabeth Lomas, School of Computing, Engineering & Information Sciences
- Professor Maia Angelova, School of Computing, Engineering & Information Sciences
- Professor Glenda Cook, School of Health, Community and Education Studies
- Professor Anthony Hildreth, School of Health, Community and Education Studies
- Dr Jeremy Ellman, School of Computing, Engineering & Information Sciences.

The internal advisory Panel comprised:
- Records and Information Manager
- Research Development Manager, Research and Business Services
- Ms Ellen Cole, Information Manager (Institutional Repository), Library
- A member of The Graduate School

The team would particularly like to acknowledge the contributions from the Expert Panel members, including their review of the draft recommendations, and the SharePoint prototype developer Mr Daniel Wainwright.

2 Project Summary

This collaborative research data management planning project (hereafter the RDMP project) sought to help a collaborative group of researchers working on an EU FP7 staff exchange project (hereafter the EU project) to define and implement good research data management practice by developing an appropriate DMP and supporting systems and evaluating their initial implementation. The aim was to 'improve practice on the ground' through more effective and appropriate systems, tools/solutions and guidance in managing research data. The EU project (MATSIQEL - (Models for Ageing and Technological Solutions For Improving and Enhancing the Quality of Life), funded under the Marie Curie International Research Staff Exchange Scheme, is accumulating expertise for the mathematical and computer modelling of ageing processes with the aim of developing models which can be implemented in technological solutions (e.g. monitors, telecare, recreational games) for improving and enhancing quality of life.[1] Marie Curie projects do not fund research per se, so the EU project has no resources to fund commercial tools for research data management. Lead by Professor Maia Angelova, School of Computing, Engineering and Information Sciences (SCEIS) at Northumbria University, it comprises six work packages involving researchers at Northumbria and in Australia, Bulgaria, Germany, Mexico and South Africa. The RDMP project focused on one of its work packages (WP4 Technological Solutions and Implementation) with some reference to another work package lead by the same person at Northumbria University (WP5 Quality of Life).

The RDMP project's innovation was less about the choice of platform/system, as it began with existing standard office technology, and more about how this can be effectively deployed in a collaborative scenario to provide a fit-for-purpose solution with useful and usable support and guidance. It built on the success of the *Datum for Health* project by taking it a stage further, moving from a solely health discipline to an interdisciplinary context of health, social care and mathematical/computer modelling, and from a Postgraduate Research Student context to an academic researcher context, with potential to reach beyond the University boundaries. In addition, since the EU project is re-using data from

elsewhere as well as creating its own data; a wide range of RDM issues were addressed. The RDMP project assessed the transferability of the DATUM materials and the tailored DATUM DMP.

The project outputs have been submitted to the DCC and are freely available on the Web (http://www.northumbria.ac.uk/datum).

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)**
Outputs:	
RDM requirements questionnaire	The questionnaire we developed and used to identify data requirements which drew on the questionnaires used in the Incremental and Sudamih projects and the DCC curation lifecycle model http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/action/outputs/
RDM requirements	Summary of RDM requirements of the EU project used as the focus for the RDM project contained in a blog post comparing supporting infrastructure http://www.datumrdm.blogspot.co.uk/2012/03/comparison-of-it-infrastructures.html
A model DMP template	Provides a revised DMP with a quick to complete front sheet; plus guidance on how to complete the DMP http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/action/outputs/
Documentation of the tailoring of the DCC's DMP	A description of how & why the DCC DMP was tailored to meet the needs of the EU project staff i.e. an audit trail between the DCC DMP, the DATUM for Health DMP and the DATUM in Action DMP http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/action/outputs/
Description of the supporting infrastructure	Description contained in the final report plus a blog post comparing supporting infrastructure http://www.datumrdm.blogspot.co.uk/2012/03/comparison-of-it-infrastructures.html
Guidance & user support documentation	Guidance & user documentation covered: RDM roles & responsibilities for a research project; information architecture guidance (viz. folder & filenaming, version control; metadata guidance); information security incl. backup & encryption guidance http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/action/outputs/
Final report	A summary of the RDMP project's methods, key findings and learning, description of RDM infrastructure developed including an initial macro level assessment of infrastructure to support RDM in one institution, with recommendations on how to support research staff to plan and implement RDM using existing tools. http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/action/outputs/
Presentation(s)	• 'DATUM in Action: Supporting researchers to plan and manage their research data' JISC Launch Event, 1-2 Dec 2011, Nottingham • 'DATUM in Action' School of Computing, Engineering & Information Sciences, Northumbria University Research Seminar, 29 Feb 2012 (postponed to May/June due to illness) • 'DATUM in Action: outputs' Meeting (disciplinary) challenges in RDM Planning, JISC Workshop, 23 Mar 2012, London
Outcomes	
Improved RDM on the ground	For the EU project team (case study): (i) a specific DMP and improved supporting infrastructure for its implementation; (ii) changes in RDM practice and improved approach for the future; (ii) a model for managing the research data in the project's other work packages For the RDMP project team: existing good RDM practice was further improved by implementing refinements based on research for this project (e.g. implementing a DMP for

	the project; refining filenaming conventions; access & security)
Developing RDM at Northumbria University	An immediate outcome was the revision (during the RDMP project) of the university's research records retention schedule to plug some significant gaps (viz. retention of the research data/records per se). Action to develop the university's RDM policy as a result of a member of the Expert Panel attending the JISC workshop on RDM policy 12/13 March, 2012. Agreement to work towards a website in the Research & Business Services site collating/linking all RDM related guidance in one place for researchers; and a recommendation to develop an online training module for RDM that all researchers/research active staff must complete
Knowledge built in the project team	• For the EU project team (case study): increased awareness and knowledge of RDM and approaches/tools to implement it • For the RDMP project team: further detailed knowledge of: RDM requirements (incl. funders' requirements and plans); clearer understanding of EU data/records retention requirements; access & security solutions; IT tools (e.g. SharePoint, Bazaar) • For the Expert Panel: increased awareness and knowledge of RDM including funders' demands and future developments (e.g. RDM policy), new networks (JISC/DCC), researchers' needs for RDM infrastructure support
Networks/ collaboration	The RDMP team consolidated and extended internal networks/collaboration within our own School, the School of Health, Community & Education Studies, Research & Business Services, the Records & Information Manager, the Institutional Repository (Library) and The Graduate School. We brought Research & Business Services and the Records & Information Manager together and they will collaborate on providing better guidance/training; and the Institutional Repository Manager is now cognisant of the potential demand and needs of researchers in terms of a data repository in addition/linked with the publications repository.
RDM case example	The final report and *DATUM in Action* project website/blog capture this case example of a straightforward, fit-for-purpose design for other researchers on similar projects to use for RDM. The collaborators plan to write an article with covers this and the outcomes of the *DATUM for Health* project for wider dissemination.

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

The RDMP project was conducted via three action research cycles. Each cycle comprised action planning, action taking phase and reflection. The action planning and taking were specific to the particular cycle but for every cycle reflection required both sets of project staff (the RDMP Project and the EU project) keeping reflective diaries in the form of posting to a RDM blog (available only to the RDMP and EU project staff). The posts were collaboratively tagged and analysed to identify themes.

Cycle 1: Requirements analysis of the EU project's RDM requirements
We sought to identify an existing tool to use to identify RDM data requirements. No single tool was ideal so we developed our own questionnaire to identify requirements which drew on the questionnaires used in the Incremental and Sudamih projects and the DCC curation lifecycle model.[2] CARDIO was considered but was not felt to be helpful at the level of a specific research project. The scene of the RDMP project was set at the first meeting and the EU project staff were introduced to RDM issues, the data curation lifecycle and data management plans. After this meeting EU researchers completed the questionnaire. Then a focus group was held to discuss these and to describe the EU project in detail. This step helped us to develop an understanding of the information and data being used and created in the EU project and helped the EU project researchers to understand the concept of the data lifecycle and a DMP since we used them as the framework for the requirements questionnaire. The outputs from Cycle 1 were the data requirements questionnaire template and a short report (in the form of a blog post) detailing the RDM requirements analysis for the EU project and a description of the wider information and data development.

Cycle 2: Development of a data management plan (DMP) for the EU project
Starting with the *DATUM for Health's* DMP (a tailored version of the DCC's DMP developed in the DATUM for Health project), a more detailed DMP was developed for the EU project customised to meet the requirements identified in Step 1. Initially EU project staff started to complete the DATUM for Health DMP for the MATSIQEL project. Their comments on this process (as both notes in the DP

itself and discussions at a meeting) made it clear that the DMP template was inadequate for the needs of a research project and therefore it was further modified (see 3.3). The way the tailoring was done was documented in the form of an audit trail between the original DCC DMP, the *DATUM for Health* DMP and the *DATUM in Action* DMP in order that the DMP and the approach can be adopted by others (at Northumbria and elsewhere). The EU project team staff were asked to complete the tailored DMP and were mentored through discussion at meetings and follow-up email queries. As they did so the guidance required to implement the DMP was identified. The aim was to develop all supporting guidance documents using a wiki on the University's VLE, available to both projects' staff, however this ultimately was not possible. The University's wiki technology does not allow for restricted access and, since the level of detail of the guidance was unknown at the start and there was some sensitivity with the EU project data, it was decided not to take this route to assure security and confidentiality. They key outputs from Cycle 2 are: (i) a DMP template; a guidance document for completing the DMP; an audit trail between the DCC DMP, the DATUM for Health DMP and the DATUM in Action DMP.

Cycle 3: Development of supporting infrastructure for RDM to implement the DMP
The choice of infrastructure was (i) existing commonly available, standard office IT software (Windows Office Suite) and a secure shared network drive and (ii) collaboration software using SharePoint. This was deliberate on the basis of familiarity and availability first and then to assess the ease of developing a SharePoint environment offering greater functionality. For (i) we developed the high-level design for the information architecture, together with support in the form of advice, guidance, templates, access protocols etc. and an implementation plan. The implementation plan involved a member of the EU project team placing copies of existing project documents/records in the file plan (folder structure) developed for the project. A decision was taken not to rename existing files (this would have taken too long) but the EU researchers committed to using filenaming guidance for all new files/documents. Our guidance drew on existing guidance and material from the university, other HEIs, the UKDA etc, tailored to be suitable for the audience i.e. researchers rather than data curation experts. For (ii) we developed an experimental prototype collaborative infrastructure environment based on SharePoint hosted by a third party provider in the cloud. The information architecture mimicked the one developed in (i) in terms of the file plan with the addition of more sophisticated version and access control and a blog and wiki. It was not possible to integrate email under the licence available or to create a public facing web site, although both are possible under a more expensive licence agreement. The EU team evaluated both infrastructure approaches. For (i) this was done first by asking for comments on the guidance, resulting in refinements as necessary, and via a focus group discussion around a series of questions; they were asked to evaluate it in terms of fitness-for-purpose, usability and transferability to other projects / departments, and to make suggestions for improvements. For (ii) it was done via a form of cognitive walkthrough with the developer demonstrating the collaborative prototype; they were asked to assess the user interface (criteria: Visually pleasing; Clear language; Clear navigation; Easy to follow/learn; Options for both new and experienced users; Clear user instructions; Help facilities; Easy to recover from user errors; Accessible from many different organisational/geographical locations; Can support large amounts of data; Training available for operators and users; Any additional comments) and to comment on the Pros / Strengths and Cons / Weaknesses of setting up key functions, i.e. Fileplan, Access rights, Version control). The RDMP staff undertook a separate cognitive walkthrough of the prototype.

An additional infrastructure tool was explored which was not in the original plan. Software version control was identified as a RDM requirement and therefore the Bazaar open source software for distributed version control was evaluated in the context of the EU project environment (http://bazaar.canonical.com/en/). A proof of concept implementation was required; this was made up of two branches: a server side implementation and one client side. The client side implementation served to evaluate Bazaar's claims of ease of use - would a Microsoft Windows user really find the software was accessible? The server side configuration was tested using a basic Amazon Machine Instance (cloud computer -- free tier). This virtual Linux system runs a command line only, but had Bazaar pre-installed.

We used the CARDIO Quick Survey [3] (9-questions) with both the EU project staff and the Expert Panel to benchmark the RDM infrastructure at the University. With the EU project staff we also used the DCC's exercise on 'Developing a roadmap for Research Data Management' to identify the top priorities for RDM.[4]

Towards the end of the *DATUM in Action* project the Expert Panel, comprising staff in relevant central service roles (viz. a Research Development Manager, the University's FoI Officer/Records Manager, a member of the Institutional Repository team), were asked to comment on the guidance produced. The university's Information Security Manager was specifically invited to comment on the information security guidance. Their comments were addressed in finalising the guidance. The draft recommendations were also discussed at the final Expert Panel meeting and refined for this report. Post the RDMP project, relevant recommendations on developing RDM will be made to the Pro Vice Chancellor (Research) and Research & Business Services Northumbria University in the form of a report.

3.3 What did you learn?

The MATSIQEL project proved to be a very good case example for this RDMP project because it combines numerous requirements and restrictions for research data that render its data management requirements realistic, rather than an academic exercise. It is using an external, real-world dynamic data set as well as research project generated data. This highlighted many practical issues e.g. data sharing/transfer, ownership and IPR, confidentiality, anonymisation, multiple versions of both raw data and derived data requiring robust version control/tracking, required/permitted access rights, geographic boundaries (EU/non-EU) etc). The research data concerned is less than 100Mb. These issues will be common to other disciplines/contexts and hence the learning has potentially much wider value. They demonstrate that the real world is messy!

We found that there is a low level of knowledge about RDM amongst staff whose research discipline is not information management/computing. Staff from a health background do, however, have a very high appreciation of the ethical issues, and the need to keep confidential data secure, and have very good procedures for doing so. Current practice does not always involve the use of a team shared drive or, where this is used, its organisation is very 'individualistic' (i.e. not based on the research processes and often adopting 'personal' filenames (e.g. joe's database). However, once research staff have engaged in the concept of good RDM they can be very enthusiastic, committed to applying good practice prospectively and cascading this good practice to colleagues.

- *RDM requirements analysis*

Research staff are not always aware/clear about their RDM requirements and can find articulating them a challenge. This can be due to lack of awareness/information about the funder's requirements, the institution's requirements and their own needs during and after the research, or the nature of the research project. In the case of the EU project the funding was for mobilities (i.e. travel to share knowledge and develop collaborative research activities) which meant that the research project activities resulting from the networking meetings could not be fully anticipated in advance and had therefore not been scoped in great detail in the MATSIQEL proposal.

No suitable existing RDM requirements analysis tool was identified. CARDIO was considered but rejected as not being appropriate at the level of a research project; it is more suitable for assessing infrastructure at the institutional level. The questionnaire we developed, based on the DCC's data curation lifecycle, proved to be useful in structuring a discussion that helped us to (a) develop an understanding of the information and data being used and created in the EU project (b) identify the RDM requirements and (c) simultaneously share the lifecycle concept of managing research data with the researchers. However, during the next step (Cycle 2: Development of a DMP for the EU project) it became clear that there was duplication in the discussion and coverage between the data requirements questionnaire and the DMP template. On reflection the requirements analysis step could have been omitted since the requirements became clear during the completion of the DMP. In this scenario it might have been better to have started with the DMP and from that extracted a summary of the RDM requirements. Data requirements analysis can be interpreted in many different ways: at the strategic/institutional level or at the operational/project level; in the context of system design/specification or systems infrastructure. For the EU project, being at the operational level and using existing infrastructure, there is no real difference between the data requirements analysis and the DMP.

In essence our analysis of the discussions with the EU project team identified their data requirements as:

1. Management of ethical approvals for the different data sets being used and generated (forms, permissions, correspondence; audit trail of approval/associated activities) and the consent for use/re-use of the datasets
2. Ensuring anonymity of data
3. Ensuring confidentiality, particularly when sharing data between EU project team members; some of the data is covered by a non-disclosure agreement
4. Version control (incl. audit trail, tracking) - multiple copies of raw data and processed data (due to data preparation/manipulation by different groups using different methods)
5. Version control of processing software
6. Distribution control – due to different research centres
7. Permissions management – within and across work packages, within/outside Northumbria University, across geographic/jurisdictional boundaries
8. Central shared data space – to ensure authoritative records
9. EU funder requirements

Since the EU project is an exploratory project in an under-researched field requirements will change. It was interesting that some of these requirements were of greater concern to some members of the team than others, though no-one disputed the range of requirements. For instance, ethical approval, confidentiality and permissions management were high on the list for the health/social care discipline researchers; version control of processing software and distribution control were high on the list of the computing / maths discipline researchers responsible for data preparation and analysis[5].

The EU funder's RDM requirements beyond the need to retain financial records were not well known to the researchers. The University's Information & Records Manager (and member of the Expert Panel) was also keen to identify a definitive source of guidance about the EU's RDM requirements. It was agreed that the RDMP project PI would follow this up because of her network of contacts. The EU advised that "there is no general EU policy for the retention of documents produced by organisations funded by the EU, as the rules of the various EU Funding Programmes are very different." This means that the preservation of documents by the beneficiary (HEI) changes according to the EU programme which finances the project. Their requirements seem to focus on the retention of financial data/records associated with the funding provided, rather than research records, and they can vary according to the particular funding programme with some programmes requiring retention of the *original* documents or *authenticated* copies.[6]

- *Fitness-for-purpose of the DCC data management plan (DMP)*

The customised data management plan (DMP), based on that of the DCC and developed in the *DATUM for Health* project, was found to be inadequate for this project. Whilst useful for raising awareness of RDM issues with the EU project team the DMP did not capture decisions and actions and therefore did not proactively support implementation. As such it was aspirational rather than actionable. It could remain a 'tick-box' exercise that was never implemented. A significant amount of work was undertaken to develop the DMP to be fit-for-purpose. This included the following design choices:

- A focus on the researcher and what is needed to help *them* conduct their research project on a day by day basis. The focus was therefore moved away from (i) data curation and (ii) sharing data after project completion. If researchers see RDM and DMPs helping *them* do their research they will use them, and from that will flow data re-use and data sharing as applicable.
- Detailed help provided in a separate document rather than within the template. For certain key topics additional guidance documents have been produced. However there is still a need for a small amount of help within the DMP itself: (i) for issues that are applicable to most research projects, specific action points have been given; (ii) examples of how to fill in particular sections have been provided.
- Addressing researchers' concerns that the DMP was duplicating work that already exists, particularly with ethical approval. Therefore, rather than requiring details to be included in the DMP the 'location' of documents providing the required information is requested. This location

could be a filing cabinet in a named office, a folder within a shared drive, a website (for external bodies) and so on.

- Cutting down repetition within the DMP itself where similar information is asked for in different sections, e.g. privacy in both the ethics section and data protection section. This has been addressed by the use of 'see' and 'see also' references.
- The previous DMP template lacked the required links to decisions and actions which is crucial if a DMP is to be helpful to the researcher and be a living document throughout the project duration. Therefore sections specifically requesting decisions/actions have been added. In some places actions have been added which are probably required by most projects to act as a guide, however it is also indicated that the researcher can add in any other actions they require (i.e. Other (specify)). Every action is accompanied by a Done box so they can be checked off once they have been completed.
- A section documenting RDM roles and responsibilities has been added near the beginning of the DMP.
- A data analysis section has been added.
- A dissemination section has been added.
- Previously separate sections have been merged so the reasons for requesting particular information become clearer. E.g. in the data collection/analysis sections the file formats are directly linked to a particular method, and the hardware/software needed to carry out that method.
- Provision of a 'DMP Lite' by listing all the sections at the start of the DMP. The researcher can then select and complete the sections that are currently applicable to a given research project, and the sections not currently applicable can be ignored. This cuts down on unnecessary work, and could be particularly helpful for small projects. Additionally, it accommodates different disciplines where some sections may never be applicable, e.g. much scientific research does not require the 'ethical issues related to research involving people' and the 'data protection' sections.
- The amended DMP template comprises tables within a Word document. Ideally a DMP needs to be an electronic tool (cf. DCC DMPOnline). This would provide: (i) the ability to easily skip currently not applicable sections; (ii) the ability to click on [i] buttons to see guidance notes, and links to bigger guidance documents; (iii) the ability to mark 'actions' with traffic lights, e.g. red = needs to be urgently addressed, yellow = needs to be addressed, green = done; (iv) connecting together sections that have to be separated in a table because of space constraints: ideally for a given research method/technique the data continuum from collection of raw data through to dissemination of outputs needs to be connected in one sequence with the RDM issues, decisions and required actions as branches from each stage in the continuum.
- The DMP needs to be integrated with other systems within the University, e.g. the JES form for proposal development, the online ethics tool, and so on.

[Note: A number of our design developments appear to be addressed in the DCC's latest version of DMPOnline (V3) demonstrated at the JISC MRD Workshop 23 March 2012.]

We learned that the type of guidance the EU project staff needed to complete and implement the DMP, beyond the support/mentoring from the RDMP project staff, was very practical guidance for RDM, specifically on roles and responsibilities, folder structure design, filenaming and version control, metadata and information security. A lot of guidance already exists both in the university and in the public domain. However, this is often either unknown to the researchers, not at the practical 'how to' level and/or not written for a researcher audience. Most of the EU project researchers wanted 'plain English' style guidance but other very technical researchers might require more sophisticated 'techie' style guidance. It is vital to understand the audience for the guidance and tailor it accordingly. Ultimately the guidance developed was not sensitive and therefore using a wiki, rather than circulating versions, might have been possible. The guidance was developed by the RDM team with only review comments from the EU project staff.

Completing the DMP was burdensome for the EU researchers. However, there were a number of other, one-off factors that contributed to this 'burden': (i) the researchers saw very little difference between the DMP and the previous data requirements stage of the project and therefore saw this as unnecessary work; (ii) the *DATUM in Health* template needed modifications; (iii) some parts of the

DMP were being completed retrospectively as the EU project had been going for some time before the *DATUM in Action* project started. The EU project researchers felt that if a DMP was completed prospectively and updated regularly the burden would be far less onerous.

Completing the DMP was beneficial to the EU project researchers: it helped raise their RDM awareness and highlighted necessary actions for their project. Benefits included: highlighting the complexity of data; realisation of how little attention is given on a daily basis to setting up systems to protect data; identification of the range of systems that exist and generally are not widely used; awareness of the different understandings of the nature of personal/sensitive/anonymised data across disciplines; realisation that some project information is only held by/known to one researcher making it vulnerable; identifying the crucial need for all members of the research team to comply with the processes that are included in the DMP.

The EU project researchers felt there was a strong case for the development of DMPs, but that a proportionate, balanced approach was needed. A DMP must not be too onerous and needs to be user-friendly to optimise compliance. However, if the DMP was merely good practice and there were no requirements for its use then the pressures of daily research activities and project deadlines could result in cutting corners or non-compliance. The DMP needs to be proportionate to the nature of the project – there is a difference between a small scale project, or a one-person project, and a major, multi-centre project with large numbers of data sets. For the first situation a 'DMP Lite' would be needed.

A major issue with the DMP is engagement of researchers. Engagement cannot only be via a stick approach, i.e. a requirement to complete a DMP. This could lead to researchers treating the DMP as a one-off, tick-box exercise. The DMP also needs to offer benefits to researchers. It needs to assist them to conduct their research project more efficiently and effectively, not be a mechanism mainly for data sharing after project completion. RDM awareness raising and training is also necessary.

We were unable to estimate costing information for RDM resource allocation. However, in this particular instance the costs were mostly staff time and a secure location for data storage and sharing by staff in different institutions worldwide. Staff time is very difficult to estimate, particularly at the start of a project such as the one here when the details of the research has not been developed, and when there is little RDM costing data to draw on for experience in making estimates. From a technology perspective the cost elements are: shared drive space; encryption facilities for laptop/usb sticks (for mobility purposes); SharePoint licences on site/third party service costs.

- *Supporting infrastructure for RDM to implement the DMP*

From the individual reflection by both the EU and RDMP project staff and the focus group evaluation of the supporting infrastructure, we learned that using existing, readily available, standard office software and shared drives, with a robust information architecture, templates, guidance etc, was fit-for-purpose for much of the EU project's RDM requirements. However, the shortcomings of this approach are that: (i) it relies upon human, manual action to establish access rights, capture metadata, control versions (through appropriate filenaming conventions), and implement retention management (at a future point); (ii) it creates problems for access by staff external to the university; (iii) it does not offer collaborative functionality. The SharePoint prototype collaborative software environment addresses all of these shortcomings, including access *if* it is implemented using a third party service/the cloud. The potential downside is the location of the hosted data; data protection legislation must be met. Being a cloud service, access to all project researchers is easily arranged by giving them ids and passwords. The number of people able to use the site is governed by the price paid. However there are issues with data protection: the country of origin of the cloud service provider needs to be in the EU or have a safe harbour agreement. How secure is the service? One would expect the cloud service provider to offer the same level of security as a University, but what access do the provider's IT staff have to the data? And how are backups handled if the provider goes out of business? A service level agreement will address some of these issues and is recommended that when using a cloud service provider. If SharePoint was made available by the University than the same external access issues would occur as with the standard office IT facilities.

The EU team welcomed both the better structured shared folder area and accompanying guidance and were mostly very positive about the collaborative prototype. Only one person was concerned about the potential 'barrier to entry' to using SharePoint.

(a) About RDM infrastructure generally:

- The need for data management processes to be an on-going, live aspect of any project that all team members buy-in to.
- The need to establish data ownership and how to address this. Does the DMP help?
- The need for RDM training and awareness raising – how can we expect researchers to implement RDM fully if they don't know what they don't know?
- The need for better support for *implementing* research data/records retention, especially in the context of the demands from funding councils for long term data/records storage.[7] Setting retention periods is not enough, support is needed for easy implementation i.e. retention or destruction at some point.
- The need for a one-stop-shop for RDM support (policy, guidance etc) since RDM guidance exists in many locations within and outside the university, meaning researchers either need to know it exists or might exist and know where to look. The most appropriate place for this is a University's central research service website (in the case of Northumbria Research & Business Services) with links to other locations (e.g. FoI/Data Protection, records management, institutional repository etc. and external bodies).[8]

(b) About standard office software plus shared drive:

- Guidance is necessary. It needs to be tailored to the research context and for both the novice to RDM and the more experienced/aware, the more/less technical person. Some guidance needs to be very specific, e.g. linked to individual tools/systems being used by the researcher. The EU project team discussed whether some guidance could be adapted into protocols/standard operating procedures (SOPs) – these are familiar approaches in the health and maths/computing disciplines. A procedural approach could aid consistency and compliance across the team.
- Although a fileplan is not 'rocket science', it is an unfamiliar concept to many researchers. It is also not as simple to produce a model fileplan for a research project as might be expected as all research projects vary in their nature, size and the demands placed upon them, and researchers' preferences/habits vary in terms of organising data/files. However, the functional approach adopted was welcomed and relatively intuitive (since it followed the research process)
- Applying access rights needs to be easy for researchers. For example, here a request had to be made to IT Services for access rights to the shared drive but subsequently seeing who had what rights was not possible from the RDMP/EU project staff's desktops due to local rules/set-up. It is possible to establish simple access rights (e.g. at the group level), to enable/disable inheritance at various levels and, exceptionally, to set access rights at the document level. But all of this is manual, requires familiarity (and therefore training) with permissions management and needs documenting.
- Lack of standard tools, e.g. automatic versioning of files, automatic retention management, minimal automatic capture of metadata. It is very difficult to change established patterns of working, particularly if the appropriate RDM tools are lacking. Even if RDM tools are made available the time to learn to use these new tools can be a barrier (once a researcher has started a new project there is no spare time for training).

(c) About the SharePoint collaborative prototype:

- Both the EU and RDMP project researchers were attracted to using the SharePoint prototype for managing research data. It met all the requirements, particularly for collaboration and data sharing. It provided much more functionality than the standard infrastructure (e.g. automatic and customisable version control; detailed access control/permissions management mostly at group rather than individual level; the potential for a retention management 'add on'). It looked attractive and easy to use.
- The benefits of SharePoint come at a cost. Use of a cloud-based service would require funding, e.g. as an item within a research proposal's budget. Universities, who have not already done so, could set up a SharePoint implementation, either across the institution or for

a specific activity such as research. Basic SharePoint comes as part of the academic campus license agreement, however, staff resources are needed to set up and run the implementation.

- The barrier to entry i.e. all researchers would need to learn how to use a new system, and how to set up a team site. But this barrier is not high. It could be likened to the adoption of VLEs within universities. Initially there was opposition from some academic staff, now all staff use them as a standard system. A VLE is set up with standard module templates; help guides and training are available. Similarly in SharePoint, a research project template for team sites could be made available which could then be customised by researchers for specific projects. However, uncontrolled use of team sites can lead to SharePoint sprawl.[9]

(d) About the Bazaar open source distributed version control software:
This additional infrastructure tool was explored because software version control was identified as a RDM requirement[10].

- In terms of the client side implementation, the Bazaar Explorer Windows client was functionally identical to that running on other systems. Once the research data was installed as a repository on a managed server, Windows configuration was simply a matter of inserting an appropriate URL and storing a user name and password. Thereafter all the Bazaar operations were integrated into Windows Explorer Context Menu (the menu that appears when a user clicks right). Users are able to download the current version, amend their own versions, and be able to store, and retrieve any incremental changes they had made. A Microsoft Windows user should find the software is accessible.
- In terms of the server side configuration, setting permissions is a System Administrator task, but did not require any unusual knowledge or skills. Bazaar actually supports several network protocols. It can be run as simple extension of an Apache Web Server, or uses a special, more efficient network protocol. For the proof of concept the Apache approach was used.
- Further user trials are required to confirm the initial very positive impression that Bazaar fulfils the EU project's RDM requirements for software version control.

(e) About the University's research data infrastructure:

- Using the CARDIO Quick Survey [11] with both the EU project staff and the Expert Panel to benchmark the RDM infrastructure at the University resulted in mostly B ratings with some A/C ratings suggesting we are 'making steady progress', have our 'finger on the pulse' in some areas (data storage) and need to 'boost' aspects of our research data infrastructure.
- No-one found the CARDIO Quick Survey easy to complete; the first question everyone asked was 'who is this aimed at" followed by 'I don't know enough to make a judgement' or 'one statement in a rating is true but not the other'. We agreed that we would complete the survey from our own perspective and perception of what we thought was true of the institution.
- In using the DCC's exercise on 'Developing a roadmap for Research Data Management'[12] with the EU project staff we only asked them to identify their priorities from those activities in the diagram. Their top three were: 1. RDM policies; 2. Funding; 3. Tools. Without a policy being funded there will be no tools; underpinning any policy and funding allocation by definition would be 'Advocacy (senior management)'.

- *Action research approach*

The variation of standard action research methodology (i.e. practical problem solving) employed was successful. It was based on the fact that the 'problem' (viz. RDM) had already been diagnosed and its participative, emergent and reflective aspects were a good fit for the rapidity of the project in developing and implementing a RDM plan and supporting infrastructure. We used a private blog to collect reflections. Though it took a little while for the EU project team to use this, as they were unfamiliar with blogs, it has proved a very useful resource for capturing reflections, lessons learnt etc. The EU project team did not use the tagging facility; if such a technique was used in future then specific training on this aspect would be needed. The RDMP team blog administrator retrospectively tagged everything to aid retrieval of information for analysis and writing outputs.

3.4 Immediate Impact

The EU project staff said that collaborating on this RDMP project had increased their awareness of RDM and helped them understand what improvements to practice could be made. Some participants felt more confident about planning and managing their research data. They began to apply the RDM knowledge and experience gained to the EU project. The WP Leader, keen to share their learning and the infrastructure developed to other EU project members, began this process at one of their workshops (14/15 March 2012) by inviting another WP leader from one of the other countries to participate in the collaborative prototype walkthrough meeting. She plans to share the RDM experience during a forthcoming exchange visit to Mexico and will definitely apply the RDM practice learned (including the tools / guidance) to her future projects and will cascade it in her own School and elsewhere (there has already been one example of this). As a result capacity building becomes sustainable (bottom-up) and the message that 'good RDM processes are required wherever 'research' takes place' is spreading.

RDMP project staff also learned new things and, although we already practiced good filenaming, we are actively implementing further refined filenaming as per the guidance we developed. But old habits die hard and not all those involved in the project have changed! For example, previous filenaming custom and practice from a software development context, where code names are the norm until a product is fully developed and launched, is being used.

By engaging the Expert Panel members in the RDMP project, deliberately chosen because of their central role in RDM related activities we have had some immediate impacts. The research records retention schedule has already been revised to cover research data[13], which is missing from the JISCInfoNet Retention Schedule[14] which many HEIs have used as their model. We alerted Research Business Services (RBS) to the JISC meeting on research data management policy (and EPSRC reminder about developing a clear RDM by 1st May 2012)[15]. As a result the RBS Expert Panel member attended the JISC meeting and has been asked to prepare a position paper on RDM in the university.

In summary the benefits have been:

- For the EU project researchers a raised awareness of the importance of RDM and of JISC's MRD programme and practical guidance for implementing RDM
- For the RDMP project team members new collaborations, including an invitation from the EU project team WP Leader to participate in a series of workshops to scope future research with a health sector organisation as a result of this collaboration.
- For Expert Panel members working in research support, FoI/records management and the University repository, an increased understanding of the importance and relevance of RDM to their activities, their role in actively supporting effective RDM and how they can collaborate in providing support; knowledge of current developments in RDM (e.g. RDM policy/roadmap) and networks
- For the wider research community, elements of a potential RDM model infrastructure by virtue of sharing the learning and guidance via the website, blog and JISC workshop *and* cascading these within the wider EU MATSIQEL project in the UK and internationally.

Evidence of this impact has come from our reflection meeting with the EU project team, from blog posts, from discussion in project meetings and elsewhere.

Some of the Expert Panel members were also involved in the *DATUM for Health* project; their continued involved in *DATUM in Action* demonstrates their commitment to improving RDM on the ground within the University.

3.5 Future Impact

The elements for a 'model' system, including a researcher focused DMP, supporting infrastructure using commonly available standard office technology and an implementation process, now exist and can be used in the future by other researchers with similar RDM requirements. The EU project team can simply share the RDM infrastructure developed with other work package team members, irrespective of their geographical location. However, if the infrastructure elements are to have wider

use and impact they need to be made readily available, visible and known to the research community. The Expert Panel discussed the notion of developing a one-stop-shop for RDM guidance, news etc, ideally located in the Research & Business Services website, which would link to advice available elsewhere in the university (e.g. in the University Secretary's Office, HR for training). A recommendation will be made to Research & Business Services regarding the development of such RDM infrastructure, both technical and support, covering policy and procedures as well as technical, human and financial resources, and to HR in terms of an online RDM training module for researchers.

The RDM team plan to use and refine the early experimental prototype SharePoint collaborative site for research projects before making recommendations about its adoption. The aim is to develop one or more 'site' templates to provide a 'vanilla' structure for research projects to customise.

Impact will be monitored via Web usage statistics of the DATUM in Action pages and ad hoc feedback and discussion with the EU researchers and the Expert Panel members.

4 Conclusions

The DATUM in Action RDMP project focused on a relatively complicated, but not atypical, real life research project. It proved to be a very good example because it used external, real-world dynamic data and research project generated data and had a range of data requirements and restrictions. This highlighted many practical issues (e.g. data sharing/transfer, ownership and IPR, confidentiality, anonymisation, multiple versions of both raw data and derived data requiring robust version control/tracking, required/permitted access rights, geographic boundaries) which are common to other disciplines/contexts. This report acts as a case study for others planning RDM for similar projects or groups, including lessons learned, guidance and recommendations (see below).

The researchers engaged in the EU project (the focus of this RDMP project) agreed that good RDM is part of the research process and therefore required wherever 'research' takes place. RDM needs to be appropriate and proportionate[16] to the nature and type of research being conducted and funder requirements. The supporting infrastructure for RDM comprises three elements: training/awareness and mentoring; guidance including policy, procedures and 'how to' guidance; technology, including software and hardware, digital and non-digital, specific tools.

Articulating RDM requirements can be difficult for researchers, either because of the nature and stage of the research project or their lack of awareness and appreciation of the scope of RDM. However, the DMP is a very useful tool in helping them to understand and articulate their RDM needs. At the level of a research project the DMP is/can be the most appropriate tool for a requirements analysis, rather than a tool such as CARDIO which is focused more at the department/institutional level. Using the DMP in this way eliminates the first step in the approach taken here and is therefore a 'win-win' situation for the researchers.

Based on our experience of significant further tailoring of the DMP template provided by the DCC (V2.2), their current DMP model requires further development. Its audience is more the data manager/scientist and less the lay person researcher; its focus is more on depositing data in a repository that will be retained permanently and therefore on the end of the data curation lifecycle. For many projects and researchers there may be little if any data that is appropriate to retain permanently; placing the data on a website may be sufficient. And researchers are focussed on conducting their research project to a high quality and within deadlines, i.e. the early stages of the data curation lifecycle. Therefore, the DMP does not engage them in RDM or help them to manage the data for themselves/the purposes of the research project. Engagement is vital to avoid a 'tick the box' approach to completing a data management plan. The DMP model does not include *action* so there is a danger that it is treated as a tick-box exercise, and remains a plan that is not implemented or not fully implemented. [Note: the DCC's DMP V3, demonstrated at the JISC MRD Workshop 23 Mar 2012, appears to address some of the development requirements highlighted above].

We adapted the DCC's DMP template for the *DATUM for Health* project and further adapted it for this RDMP project, focussing on the researcher and the efficient conduct of the research project. This DMP would benefit from trialling with researchers in other disciplines. However, it is a Word document and would need to be turned into an electronic tool to be fully helpful to researchers.

Three important aspects of RDM infrastructure are: human, procedures/practice and technology. Human infrastructure supports *engagement* e.g. through awareness raising, mentoring and advice. Procedural infrastructure enables researchers to '*do RDM themselves and with confidence'* e.g. via templates, exemplars, best practice guidance. Technology infrastructure provides the platform(s) and further tools to *implement* best practice RDM. The technology infrastructure does not necessarily need to be sophisticated. A fit-for-purpose technology infrastructure may be the appropriate use of standard (lowest common denominator) office software and shared drives. However, these need to be organised, used well and consistently, and require manual (human) management of for instance access rights, retention management, metadata. For collaborative purposes, rather than mostly sharing/access, then a collaborative environment such as SharePoint or GoogleDocs may be the fit-for-purpose solution, providing more sophisticated, automatic management of versioning, metadata and retention. However, depending on the researchers' skills set and/or the implementation of software such as SharePoint within an HEI, the entry to barrier or effort required to use them may be too great, not represent a good return on investment. For version control in contexts where there is a data pipeline and multiple data users then alternative, more sophisticated/flexible tools exist and are open source.

It is only when fit-for-purpose tools, systems and supporting infrastructure are in place and adopted by researchers that RDM will be successfully embedded in research culture, ethos and practice. However, *the real world of research is messy;* therefore, a pragmatic and proportionate approach is required.

5 Recommendations

The following recommendations are made to support staff to *plan and implement the management* of research data in the context of using existing tools viz. Windows operating system, shared drives and collaborative tools such as SharePoint.

5.1 General recommendations to HEIs

1. HEIs should adopt/promulgate a pragmatic, proportionate and risk-based approach to RDM in the real world of research which is *messy*.
2. HEIs should develop a RDM strategy and action plan for its implementation to support researchers on the ground (cf. EPSRC roadmap requirement[17])
3. HEIs should strategically identify projects and researchers to engage in implementing RDM practice to successfully and more rapidly embed RDM in the research culture, ethos and practice. They will then cascade good practice to research colleagues and students.
4. HEIs should identify different routes/mechanisms for cascading awareness of, and training for researchers about, RDM e.g. in sessions on writing bids by considering funders' requirements for DMPs and/or data/records retention; training research support staff and repository staff who in the future could then provide support to researchers.
5. HEIs should adopt a prospective approach to RDM rather than a retrospective approach which is simply impractical.
6. HEIs (or JISC) should produce a short (2-page) layperson's guide to RDM for Principal Investigators. This should highlight drivers (sticks) and benefits (carrots) and what key issues they need to think about.
7. HEIs should recommend that a Data Management Plan (DMP) is an explicit requirement for funded research projects and align this with their research project management system. They should provide guidance for research staff on completing a DMP, including exemplar DMPs.
8. Standard software interfaces (e.g. Windows, and shared drives) may represent fit-for-purpose supporting infrastructure for managing the data of much research that is conducted HEIs should enable research teams (of more than one person) to readily establish a shared drive with a functionally based folder structure to support retention management, the ability to locally set access controls. This should be supported with guidance on good practice folder/file naming, including version control, and security. A standard folder structure should be made available to copy and amend to suit particular needs of the specific research project.

9. HEIs / researchers should assess whether or not investing in a sophisticated system such as SharePoint is necessary. Are the benefits of adopting it for a research project great enough to outweigh the costs (financial, training, development) if the system is not already implemented?
10. HEIs should provide researcher specific Freedom of Information training to try to address the concerns of staff in relation to what information they may be required to provide upon receipt of requests and clearly define what exemptions exist so as to ensure that FOI is not seen as a deterrent to sufficient record keeping.
11. HEIs should provide better data/records appraisal, retention and destruction guidance. The retention of research data per se is missing from the JISCInfoNet Retention Schedule[18] which many HEIs have used as their model, and therefore can be missing from HEI Research Retention Schedules.
12. The scope and relationship of RDM to, for example, data protection, ethics, IT security etc, which are often located in different parts of an HEI make for a wide range of sources and locations of information/guidance for researchers. HEIs should create a RDM specific area within the research support section of their institutional website, since this is the likely first point of reference for researchers, and link to the wider range of relevant guidance located elsewhere. This will provide a one-stop-shop for all guidance, news etc on RDM, which with the addition of an RSS feed, would keep research staff up-to-date on any important changes.
13. HEIS should audit and/or monitor that RDM is being carried out. At the institutional level this could be covered in research ethics/governance auditing process. At research project level the PI should be responsible for monitoring.

5.2 Recommendations for the wider community

To Research Funders

1. Funders should accept that retrospective RDM is impractical and be supportive of researchers in moving towards better RDM whilst not necessarily meeting all of the funders' requirements.
2. Funders (and other researchers/the research community) should acknowledge that decisions about what data is appropriate to retain are the researcher's, working in the context of legislative requirements, rather than the funder's.
3. Funders should agree a standard scope (i.e. areas covered) in a DMP so that at the macro level a DMP template is the same for all research council funded projects, even if the micro-level detail varies.
4. Funders should establish clear, justified guidance on appraisal and a retention schedule
5. Funders should explore other mechanisms for making research data widely available, e.g. promoting enhanced publications, and data papers: 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. ***A recommendation made from the DATUM for Health project is reiterated here.*** 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 DCC

1. The DCC should maintain a resource of funders' requirements on their website: these funders are also covered in their online DMP tool. This resource needs to be kept up to date and extended to include other funders' requirements even if they are not in the online tool
2. DCC should develop a 'DMP Lite' online template (along the lines of the one developed for this project). Researchers want a DMP to be helpful and not a burden or bureaucratic

5.3 Recommendations for JISC

14. JISC (or HEIs) should produce a short (2-page) layperson's guide to RDM for Principal Investigators. This should highlight drivers (sticks) and benefits (carrots) and what key issues they need to think about.
1. JISC should provide an equivalent of the Dropbox and Spideroak 'cloud' data sharing services so that there are no ownership and/or security concerns for researchers
2. JISC should explore the retention of research data with the EU since requirements appear only to cover retention of financial research records and vary according to the funding programme[19]
3. JISC should produce a detailed evaluation of open access tools such as GoogleDocs, Dropbox, SpiderOak etc for research data management covering access, security, data protection issues (cf work being done by members of the HE and FE Records Management and Information Compliance group[20])
4. JISC should influence/lobby Vitae to include more about RDM in their *Researcher Development Framework*[21] for PGR students, some of whom will work as part of a research project team and many of whom will continue a research career.

6 Implications for the future

The first implication is the sustainability of RDM at the research project level. The immediate project outputs (e.g. requirements analysis, research data architecture, processes, templates, policy, guidance) are sustainable because they use standard university desktop IT infrastructure. Any recommendation to use the experimental prototype collaborative software system (e.g. SharePoint, Google Docs) would need to outline sustainability issues. Since the management of research data begins with the researchers, by participating in this project members of the EU project team gained a deeper understanding of RDM and skills which they can share with other team members during their exchange visits; they will also be able to apply their extended knowledge and skills to other projects and cascade best practice RDM to research colleagues and students they supervise. However, this is an 'isolated' case made possible by the JISC project funding. Sustaining the human element of RDM infrastructure, i.e. the awareness, knowledge and skills, is vital to achieve widespread improvement in RDM on the ground. From the *DATUM in Health* project awareness and training have been embedded for PGR students; this needs to be extended to cover researchers. One option is a RDM training module which is mandatory for existing and new research staff.

The corollary to this 'bottom up' implication is a 'top down' implication viz. the need for a university RDM strategy (roadmap) as, for example, being required by EPSRC.[22]

Another implication is that systems for research data management (e.g. data management planning), ethics, bidding, project management etc are all important and in some cases share information about managing research data. They therefore need to be linked and appropriate process maps/workflows developed to minimise duplication and maximise sharing and efficiency.

Freedom of Information legislation has significant implications for access to research data and, therefore, for individual researchers and HEIs. Research staff need specific FoI training to ensure their understanding of the implications, address any concerns, advise them on appropriate research data practice (e.g. data retention, consent etc) and enable them to recognise a request and know how to deal with it.[23] Hence, recommendation 10 (above) to HEIs.

RCUKs are making demands for long term (in some cases potentially *very* long term) retention of research data. However, they represent a 'blanket' approach with seemingly arbitrary retention periods. The fundamental question is 'what data needs to be kept?' There is a danger of creating a sense that *all* research data must be kept and, by definition, shared for potential use/re-use. Is this true? Is it appropriate? A data continuum exists – from raw data to transcribed data to anonymised data to summarised or analysed data to synthesised data. Different disciplines have different needs/philosophies. Some are only comfortable with or only able to retain data from certain stages (e.g. anonymised data). Others would need to keep the raw data along with all the versions of the software used to analyse the data. Additionally, the long-term and wider value of data will depend on the nature and topic of the research project. Funders (and other researchers/the research community) should acknowledge that decisions about what data is appropriate to retain are the researcher's,

working in the context of legislative requirements, rather than the funder's. Assuming appropriate decisions have been made about what data is (potentially) valuable to keep/required to be kept, then a fundamental question is 'where should the data be kept?' Who will consider what is the best location? There are several options:

i. Discipline specific national data centres (e.g. the UKDA) for the permanent preservation of significant data 'sets'; these may be significant because of their size or their subject/nature, for instance it may be that the data could not be collected again as it relates to a particular point in time or a particular event.
ii. HEI open access repositories, either standalone data repositories or hybrid data/publication repositories. These would be useful for PhD data and data supporting funders' requirements (e.g. to satisfy the 10 year retention requirement of some RCUKs).
iii. Open access extended publications and/or research data papers. The benefits and the increased exposure of the research/researcher(s) through publication but this requires a wider acceptance in some disciplines of the concept of the 'data paper'.

For each option there is a serious question, and significant implication, about how this will and can be funded.

The *DATUM in Action* project has already resulted in developments and suggestions for further development at Northumbria University (e.g. researcher RDM practice; a prototype collaborative environment which RDMP project staff are planning to use further; RDM roadmap development; suggestion for an online RDM training module). The RDMP project team plans to write an internal report making recommendations on RDM at the university for the PVC (Research) and Research & Business Services. With EU project team members they also plan to write a journal article to share their learning from this project and the previous *DATUM for Health* project to a wider audience.

Long term project contacts:
Professor Julie McLeod, School of Computing, Engineering & Information Sciences
julie.mcleod@northumbria.ac.uk

Project outputs are available via the project web site http://www.northumbria.ac.uk/datum and via the DCC. The RDMP project website will be maintained as part of Northumbria University's website and it has been offered to the British Library UK Web Archive[3].

7 References

See endnotes

8 Appendices

8.1 Description of RDM supporting infrastructure

To support the implementation of the RDM using standard office software and a shared drive the following infrastructure was developed:

- A DMP plan, tailored to meet the needs of research projects, with guidance to help researchers complete it
- Guidance on RDM roles and responsibilities to enable PIs/researchers make decisions about who is responsible for what
- A model fileplan for structuring a shared drive for a research project with the rationale for the design to enable researchers to effectively tailor it to meet the particularities of a specific project
- Guidance on best practice folder and filenaming, including version control
- Guidance on the creation and capture of metadata about the research project and the data files/records

- Guidance on information security

All guidance contains links to other relevant University guidance/contacts and selected guidance/support published elsewhere (e.g. by RCUKs, UKDA)

8.2 RDM supporting infrastructure evaluation tools

(i) Standard office software plus shared drive evaluation

Questions used in the focus group discussion to evaluate this approach with prompts if required:

MRD is something you do for yourselves but…

1. Is formal / overt RDM useful for researchers?
   especially in the context of working with others, not lone researcher
   Does RDM have any benefits? What are they? Do you feel it is important? [why/why not]

2. What would make researchers undertake formal / overt RDM?
   What would it take/is needed for researchers to adopt and embed RDM into the research culture, ethos & practice?
   Would the sticks really work? What are the carrots?

3. What support & guidance, tools, infrastructure do researchers need?
   4 approaches to RDM support are: guidance & training; DMP Online; embedding RDM into bid writing; consultation with an expert eg a DM coordinator)
   Have we provided enough
   We discussed SOPs – this would make RDM very procedural – is this something you would still like? If so is there any value in taking some of the 'guidance' (eg folder/file naming) and making choices/decisions and then creating an SOP(s)?
   How fit-for-purpose/useful/usable are the tools, systems and supporting infrastructure we've developed? What improvements are needed?
   What is missing in terms of tools, systems, infrastructure for RDM?
   Is RDM something you want to understand more about/become more familiar with/confident about? Would you then find it easier?
   We've amended the DMP but think it is still not fit for purpose. What do you think of the principle of a DMP process? Was the DMP guidance helpful?

4. Future RDM practice
   Have there been any changes in your thinking/practice as a result of this RDM project?
   What ideas, practice, if any, would you take into your next research project i.e. continue doing?
   And…How will you go about that?
   What would you tell others to do in future?

(ii) Collaborative Infrastructure Prototype evaluation Part 1 – User Interface

Using the 1 – 4 scale below please tick the number which corresponds to your point of view for each feature/aspect of the user interface (i.e. the ease with which a user can employ the prototype to achieve their research data management goal).

1. Excellent **2. Good** **3. Not good** **4. Poor**

	⟷ excellent poor			
FEATURE / ASPECT	**1**	**2**	**3**	**4**
Visually pleasing				
Clear language				
Clear navigation				
Easy to follow/learn (i.e. intuitive)				
Options for both new and experienced users				
Clear user instructions				
Help facilities (e.g. how easy to find, how comprehensive, etc.)				
Easy to recover from user errors *(e.g. back functions in a software tool)*				
Accessible from many different organisational/geographical locations				
Can support large amounts of data				
Training available for operators and users *(e.g. embedded in the software as tutorials or case examples)*				
Any additional comments on the user interface				

Would you be attracted to using this for managing research data? Yes / No

Why / Why not?

Collaborative Infrastructure Prototype evaluation Part 2 – Ease of use/usability

The focus of this part of the evaluation is on views about **how easy setting up / amending** a SharePoint environment is as a potential creator. For each aspect / task please make notes in both columns (pros/cons).

	Pros / Strengths	**Cons / Weaknesses**
Fileplan		
Access rights		
Version control (e.g. automatic version control of files – options)		
Wiki for collaborative document development		

Would you feel attracted to setting up a Share Point Team site for research purposes? Yes/No

Why/Why not?

[1] *MATSIQEL: Models for ageing and technological solutions for improving and enhancing the quality of life*. FP7-People-2009-IRSES Marie Curie Actions. International Research Staff Exchange Scheme http://lib.bioinfo.pl/projects/view/24148
[2] Incremental http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_170910.pdf; Sudamih http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20Researcher%20Requirements%20Report.pdfprojects; DCC curation lifecycle model http://www.dcc.ac.uk/resources/curation-lifecycle-model
[3] CARDIO Quick Survey http://cardio.dcc.ac.uk/quiz/
[4] DCC's exercise 'Developing a roadmap for research data management. http://www.dcc.ac.uk/webfm_send/771
[5] See Data requirements blog post for full details http://www.datumrdm.blogspot.co.uk/2011/12/data-requirements.html
[6] See EU data/records retention requirements blog post http://www.datumrdm.blogspot.co.uk/2012/03/eu-datarecords-retention-requirements.html
[7] See for example EPSRC's 'Policy framework on research data' – Expectation (vii) for data to be "securely preserved for a minimum of 10-years from the date that any researcher 'privileged access' period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party", which could in effect mean in perpetuity http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx
[8] This finding fits with the work of DMSPpsych http://www.shef.ac.uk/[psychology/research/groups/dmsppsych/onstop
[9] Lappin, J & McLeod, J. (2010). *Investigation into the use of Microsoft SharePoint in Higher Education Institutions: Final Report*. Northumbria University http://www.northumbria.ac.uk/static/5007/ceispdf/SPfinal.pdf
[10] The demonstration of this proof of concept is available in a video available at http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/action/outputs/
[11] CARDIO Quick Survey http://cardio.dcc.ac.uk/quiz/
[12] DCC's exercise 'Developing a roadmap for research data management. http://www.dcc.ac.uk/webfm_send/771
[13] Northumbria University. (2011). Research records retention schedule http://www.northumbria.ac.uk/static/5007/uso/ResRet.pdf
[14] JISCinfoNet (2007). *HE Business Classification Scheme and Records Retention Schedule* http://www.jiscinfonet.ac.uk/partnerships/records-retention-he/
[15] Tedds, J. (2012). Chatham House at Weetwood Hall: emerging themes from the JISCMRD02 institutional RDM policy workshop. http://mrdevidence.jiscinvolve.org/wp/
[16] References at 23 Mar JISC meeting to the same concept (of a good enough solution) were General GS Patton "A good battle plan that you act on today can be better than a perfect one tomorrow." or "I would rather have a good plan today than a perfect plan two weeks from now" and Voltaire's "The best is the enemy of the good".
[17] EPSRC Policy Framework on Research Data and requirement for a roadmap by 1/5/2012 to be implemented by 1/5/2015 http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx
[18] JISCinfoNet (2007). *HE Business Classification Scheme and Records Retention Schedule* http://www.jiscinfonet.ac.uk/partnerships/records-retention-he/
[19] See EU data/records retention requirements blog post http://www.datumrdm.blogspot.co.uk/2012/03/eu-datarecords-retention-requirements.html
[20] Our Information & Records Manager advised that work is being done by members of the HE and FE Records Management and Information Compliance list HEFE-INFOCOMPLIANCE-RECORDSMGT@JISCMAIL.AC.UK
[21] Vitae *Researcher Development Framework* http://www.vitae.ac.uk/researchers/428241/Researcher-Development-Framework.html
[22] EPSRC http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx
[23] See the recent item by Adam Tickell, PVC (Research & Knowledge Transfer), University of Birmingham in light of a rejected amendment to the Protection of Freedoms Bill which would have provided limited exemptions on research data for universities. He says that "universities have been slow to appreciate the potential risks in the [Freedom of Information] Act." Since the new Protection of Freedoms Bill "will add a new requirement to permit re-use of the data" …"universities will need to read the guidance and ensure that they have defences in place to protect researchers." 'Testing the limits of freedom', *Research Fortnight*, 29 Feb 2012 http://www.researchresearch.com/index.php?option=com_news&template=rr_2col&view=article&articleId=1167097