



JISC Final Report

Project Information			
Project Identifier	<i>To be completed by JISC</i>		
Project Title	iridium - Institutional Research Data Management at Newcastle University		
Project Hashtag	#iridiummrd		
Start Date	01 October 2011	End Date	30 June 2013
Lead Institution	Newcastle University		
Project Director	Janet Wheeler		
Project Manager	Lindsay Wood		
Contact email	janet.wheeler@ncl.ac.uk		
Partner Institutions	<i>Not applicable</i>		
Project Web URL	http://research.ncl.ac.uk/iridium/		
Programme Name	<i>Jisc Managing Research Data</i>		
Programme Manager	Dr Simon Hodson		

Document Information			
Author(s)	Lindsay Wood, Victor Ottaway, Megan Quentin-Baxter, Janet Wheeler		
Project Role(s)	Project Manager, Project Management Team, Project Management Team and WP6 Workpackage Leader, Project Director		
Date	12 July 2013	Filename	iridium_final_report_v2.pdf
URL	http://research.ncl.ac.uk/iridium/outputs/		
Access	This report is for general dissemination		

Document History			
Version	Date	Comments	
1	14 June 13	First draft	
2	12 July 13	Section 2 <i>Project Summary</i> added Section 3.1 <i>Project Outputs and Outcomes</i> edited Section 6 <i>Implications for the future</i> added Appendix 8.1 <i>Evaluation of external RDM tools and training</i> added Appendix 8.2 <i>Research Data Catalogue user testing</i> added Appendix 8.5 <i>DCC DMPonline (v3) tool evaluation</i> added <i>Research data management plan template and guidance</i> and <i>Human Factors Integration mapping</i> , which were to be included as appendices are now referenced from Section 3.1.	

Table of Contents

1	ACKNOWLEDGEMENTS	3
2	PROJECT SUMMARY	3
3	MAIN BODY OF REPORT.....	4
3.1	PROJECT OUTPUTS AND OUTCOMES	4
3.2	HOW DID YOU GO ABOUT ACHIEVING YOUR OUTPUTS / OUTCOMES?	5
3.2.1	<i>Project setup</i>	5
3.2.2	<i>Conducting the surveys.....</i>	6
3.2.3	<i>Formulating the policy</i>	7
3.2.4	<i>Tools and systems</i>	8
3.2.5	<i>Support</i>	9
3.2.6	<i>Dissemination and engagement</i>	9
3.3	WHAT DID YOU LEARN?	10
3.4	IMMEDIATE IMPACT	11
3.5	FUTURE IMPACT	12
4	CONCLUSIONS	12
5	RECOMMENDATIONS	13
6	IMPLICATIONS FOR THE FUTURE.....	15
7	REFERENCES	15
8	APPENDICES	16
8.1	EVALUATION OF EXTERNAL RDM TOOLS AND TRAINING.....	16
8.2	RESEARCH DATA CATALOGUE USER TESTING.....	24
8.3	SWORD EVALUATION	25
8.4	CKAN API	25
8.5	DCC DMPONLINE (v3) TOOL EVALUATION.....	26
8.5.1	<i>Background.....</i>	26
8.5.2	<i>Findings summary.....</i>	26
8.5.3	<i>DMPonline recommendations</i>	27
8.5.4	<i>Wider institutional data management planning implications</i>	28

1 Acknowledgements

iridium was funded by Jisc under the Managing Research Data Programme. It represents a collaboration between the Digital Institute, Information Systems & Services, Medical Sciences Education Development, Research & Enterprise Services and the University Library.

We gratefully acknowledge the support of Jisc, Newcastle University, our local research community and the many individuals who have contributed to the project.

The steering group chaired by Professor Tony Roskilly, Dean of Research, Faculty of Science, Agriculture and Engineering:

Wayne Connolly (Librarian), Professor Megan Quentin-Baxter (School of Medical Sciences Education Development), Dr Douglas Robertson (Director, Research and Enterprise Services), Professor Paul Watson (Director, Digital Institute), Steve Williams (Director, Information Systems & Services).

The project team:

Ben Allen, Peter Dinsdale, Clive Gerrard, Jill Golightly, Sean Gray, Paul Haldane, Suzanne Hardy, Pete Harvey, Phil Heslop, Simon Kometa, Andrew Martin, Paul McDermott, Stephen McGough, Derek Mortimer, Niall O'Loughlin, Victor Ottaway, Megan Quentin-Baxter, James Pocock, Paul Thompson, Janet Wheeler, John Williams, Dave Wolfendale, Lindsay Wood, Simon Woodman.

aided and abetted by:

Will Allen, Rachel Armstrong, Ursula Armstrong, Doreen Boddy, Steve Boneham, Janice Coulson, Sharon Gilmore, Dave Hartland, Hugo Hiden, David Hill, Caroline Ingram, Professor Aad van Moorsel, Pete Wheldon.

The postgraduate support team:

Natalie Cresswell, Blanca Garcia Navarrete, Jack Jago, Amy-Jane Lively, Kelechi Njoku, Catherine Pruitt Velasquez, James Turland, Sathish Sankar Pandi

For JISC:

Our programme manager, Simon Hodson and Martin Donnelly, Kerry Miller and Angus Whyte of the DCC.

Finally, thanks go to fellow projects in the Programme for sharing outputs and to the Digital Curation Centre¹ for their assistance. In particular we would like to acknowledge ADMIRE² (University of Nottingham), data.bris³ (University of Bristol) and Research360⁴ (University of Bath).

2 Project Summary

The aim of *iridium* was to produce a pilot infrastructure for Research Data Management at Newcastle University. This would be achieved by scoping requirements, formulating a draft policy, providing tools to support that policy and providing a framework to support researchers in the use of the policy and the tools. A further outcome would be a business case for the implementation of the support necessary to meet key institutional data curation and research lifecycle management objectives.

Requirements were scoped via a '10 minute' online survey and an inductive thematic analysis based on interviews with researchers. Based on this, and on a review of existing relevant University policy and guidance, the University Research Office formulated 10 high level policy principles supported by a code of good practice.

On the technical side, a prototype research data catalogue was produced in order to satisfy funder requirements for discoverability of research data. This was supplemented by an investigation of

¹ <http://www.dcc.ac.uk/>

² <http://admire.jiscinvolve.org/wp/>

³ <https://data.blogs.illrt.org/>

⁴ <http://blogs.bath.ac.uk/research360/>

available tools and technologies to support RDM, the most promising of which was the CKAN open source data portal.

In terms of support, a web site was developed to provide an initial focus and to give a context to the policy principles. In addition an invitation to tender was issued for the production of training materials based on stakeholder mapping.

In the final analysis, the size of the task facing the project was underestimated at the outset. However, valuable lessons have been learned, knowledge has been gained, necessary cross-service alliances have been forged and the University is now in a much better position to take RDM forward.

3 Main Body of Report

3.1 Project Outputs and Outcomes

All outputs are available from <http://research.ncl.ac.uk/iridium/outputs/> unless otherwise indicated.

Output / Outcome Type	Brief Description and URLs (where applicable)
Report	Online survey as to the current position based on input from 128 respondents. Full and summary reports are available.
Report	Inductive thematic analysis based on 29 x 1 hour interviews with researchers. Full and summary reports are available.
Evaluation	Extensive information gathering to forming a knowledge base of external RDM tools and support/training materials utility (see Appendix 8.1).
Policy	Draft RDM policy and code of good practice as brought before University Research Committee on 12 Dec 2012.
Software prototype	Specification for a Research Data Catalogue which links data sources to publications, research projects and funders. See Appendix 8.2 for user testing.
Report	Investigation of the use of the SWORD protocol to provide easy data deposit. See Appendix 8.3
Report	Description of the implementation of SWORD endpoint within e-science central
Code	Code for the implementation of SWORD endpoint within e-science central. http://sourceforge.net/p/esciencecentral/code/HEAD/tree/trunk/code/server/SwordAPI/ Test code: http://sourceforge.net/p/esciencecentral/code/HEAD/tree/trunk/code/tests/sword-tests/
Report	Investigation of CKAN API. See Appendix 8.4
Code	CKAN Java client code https://github.com/andmar8/CKAN-Java-Client
Report	CKAN case study. Covers implementation, Shibboleth integration, data harvesting and automatic metadata attachment.
Support	The website http://research.ncl.ac.uk/rdm/ provides an initial focus for RDM support giving context to policy principles and expanding code of good practice.
Report	Human Factors Integration Mapping – associating stakeholders with support requirements
Support	Tutorial and workshop content. See Section 3.2.5.
Support	Research data management plan template and guidance.
Report	DCC DMPonline (v3) tool evaluation. See Appendix 8.5.
Business case assessment	This was brought before the University Research Committee on 30 April 2013 and proposed the recruitment of a research data manager to coordinate continuation of the work done by iridium. Available on request to janet.wheeler@ncl.ac.uk

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

The aim of *iridium* was to produce a complete holistic plan and infrastructure for RDM at Newcastle University, making data generated by research at the University both available and discoverable with effective curation throughout the full data lifecycle in consultation with the researchers. The project's methodology was based around the specific objectives that needed to be fulfilled to achieve this – in outline as follows.

- Gain a full understanding of the needs and requirements for Research Data Management by conducting a survey.
- Use the outcomes of the survey to inform the production of a policy framework.
- Identify tools and systems, including existing institutional ones, that could be used to support the requirements gathered by the survey and those generated by the policy.
- Produce support materials.
- Combine tools, systems and support materials to produce a pilot Research Data Management Infrastructure.

The aim and objectives did not change during the course of the project.

Evaluation of stakeholders (stakeholder survey) established a base line of current practice, and in addition to the topics discussed during team meetings the support team also evaluated:

- External sites, existing documentation and tools
- Internal data flows and current processes
- Deliverables developed during the project including human factors support such as the website

Costs were kept to a minimum by using survey monkey rather than focus groups and met by the institutional contribution, *iridium* support team and project manager. The evaluation criteria outlined in the project plan proved ambitious in the timeframe of the project although still relevant as RDM matures at Newcastle.

3.2.1 Project setup

A Project Manager led the project with support from a Project Management team and administrative support based at MEDEV, SMSSED. This was overseen by the Project Director. The Project Management team met every 2 weeks to review workpackage progress and Programme reporting. Project setup was initiated by establishing a project mailing list⁵, website⁶ and online collaborative environment⁷. The final project plan was submitted and required no further changes.

The project assembled a Steering Group comprised of directors of services for the project partners. The Steering Group met five times during project duration.

The project team met for an hour every 2 weeks throughout the project. Smaller working groups on requirements gathering, policy, technical tools, user-testing and business case development were convened as necessary to progress workpackages.

The project support team of postgraduate students were recruited through the University Careers Service advertisement⁸. These positions were part-time for up to 8 hours per week (dependent on Faculty guidelines) for the duration of project. Not all students who were offered a position accepted, leading to a requirement for an addition recruitment cycle. This seemed a particular issue in Faculty of Medical Sciences (FMS) where there was a suggestion that the role was not so compatible with local working practices. Overall 8 students were in post from across all three Faculties (with 1 from FMS).

The members of the postgraduate support team were inducted in late January 2012 with a session on project context and RDM basics. They were fully embedded in the project and their role constituted a wide of activities. They conducted information gathering and numerous evaluations of external RDM

⁵ *iridium* project mailing list (iridium@ncl.ac.uk).

⁶ <http://research.ncl.ac.uk/iridium/>

⁷ Project collaborative environment (<http://researchtools.ncl.ac.uk/>).

⁸ <http://research.ncl.ac.uk/iridium/vacancies/>

tools and support materials. Internally they supported requirements gathering interviews, interim analysis, policy mapping, tools and support material testing, documentation writing and website editing, blog posts and dissemination materials (see 3.1). The support team also undertook general administration duties such as arranging interviews, assisting team meetings and summarising information for reporting.

The recruitment of a technical developer within the Digital Institute was delayed as a result of an institutional freeze on appointing IT posts. This meant that the initial requirements gathering phase was delayed against the original workplan.

Internal and external stakeholders were kept abreast of developments by the project team as outlined in the dissemination plan. Formal reporting was through monthly reports via email to the Programme manager and blog posts to the Programme were conducted. Internal reporting was to the Steering Group and the University Research Committee.

3.2.2 Conducting the surveys

RDM requirements gathering was conducted by online and in-person methods. A working group led by the DI was established to formulate qualitative interview questions and one led by RES to write the quantitative web based survey. Questions were tested and modified based on pilot testing.

A wide range of research and related staff were invited to be interviewed across all the Faculties Academic Units and senior managers. Participants included research deans, principal investigators, research associates, computing support officers, data managers, technical support staff and postgraduate research students. Face-to-face semi-structured interviews of 1 hour duration were conducted with 29 members of the local research community (for template development see blog⁹) were transcribed and a two stage thematic analysis was conducted as described by Braun and Clarke¹⁰.

Full methods are described in the analysis report¹¹. In summary, the first stage of analysis was deductive, and conducted by interviewers on the transcripts of the interviews. Five initial themes were perception (concepts of data and data management), purpose (data usage and destination), process (data lifecycle), people (data lifecycle), and provoking (catch-all to gather any other salient points expressed by interviewees). The second phase used results of the deductive analysis as a starting point and attempted to build meaningful themes from whole data corpus. Analysis was done by a single researcher and generated a new set of themes of diversity, data analysis, longevity/lifecycle, responsibility and sharing and collaboration:

- Diversity: there is a great deal of diversity amongst users and any policy should enable users to achieve best practice rather than apply a one size fits all "solution" to data management.
- Data Analysis: much of the processing of data that currently takes place on local machines would be more efficiently accomplished on larger scale servers, but that users are largely unaware that such services may exist in the University.
- Longevity / Life Cycle: a strong consensus that data should never be thrown away. There should be separate systems for archiving data and current data.
- Responsibility: who should be doing what with the data, with storage, with security and with access. Many interviewees were unclear about what falls to them and what the responsibility of the University is.
- Sharing and collaboration: the concept of data access, in particular sharing data with collaborators, both internal and external, was a strong theme. This becomes problematic with very large data sets, or with collaborators insisting on using "their" systems.

⁹ Survey questions
http://research.ncl.ac.uk/media/sites/researchwebsites/iridium/iridium_main_survey_questions_for_website_16_2_2012_v1_LW.docx

¹⁰ Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. doi:10.1191/1478088706qp0630a

¹¹ Interview analysis
http://research.ncl.ac.uk/media/sites/researchwebsites/iridium/iridium_interview_thematic_analysis_5_7_2012_v1_PH.pdf

Overall from the interviews and their analysis it can be concluded that requirements were that:

- University should state minimum RDM standards expected and provide basic RDM guidance
- An integrated institutional RDM approach needed
- Researcher RDM flexibility should be supported if in line with good practice
- Ease of integration of future institutional systems/tools and policies, with existing local/research group tools/systems/policies is required
- Better promotion/awareness of RDM-related operating standards, tools and expertise together with available national services is needed
- Guidance on managing the archiving of research data would be particularly welcome
- Tools/guidance facilitating individual research group data sharing/collaboration (internally within and external to institution) are needed
- Tools for facilitating the discovery of research data as research outputs alongside web profiles/publications would be welcome

The ten minute online survey (see blog for template questions¹²) was open for approximately 7 weeks from 23 March to 11 May 2012. Invites were distributed to circa 850 research project Principal Investigators of recently active projects identified from MyProjects database to complete on behalf of research projects. It was also publicised through the Registrar's regular consultation emails to Head of Departments and through the University wide 'NU Connect' newsletter.

Responses were received from 128 research projects and full report is available from the project website¹³. A summary of findings is described below:

- 23% of research projects had a formal data management plan
- 64% of projects' data location was serviced by the institution
- 50% of research projects shared data externally
- 73% of projects shared data internally
- Data retention up to 10 years was most common
- Only 5 of 128 projects said they were aware of training sessions and materials on RDM

Further guidance was most frequently requested on the Data Protection Act, Freedom of Information Act, data security, Funders minimum requirements, NHS requirement, University requirements and data management tools. Projects were unclear on their intellectual property rights.

3.2.3 Formulating the policy

Policy analysis was an important aspect of the project led by the Research Office but involving staff from across the project team.

The work began with a thorough review of all existing relevant University policies and guidance¹⁴. Existing guidance was then pulled together with new material (identified as necessary in the surveys) to form the first content drafts. The draft policy principles were structured around the DCC research data lifecycle and the code of practice designed to map to the principles.

The draft policy principles are comprised of 10 general high level items, this is supported by the much larger code of good practice (which is in turn supported by the website). The policy principles are static but the code of practice allows for quick revision based on changes to received good practice.

¹² Stakeholder survey <http://iridiummrdrd.wordpress.com/2012/05/22/iridium-research-data-management-requirements-online-survey/>

¹³ Survey report
http://research.ncl.ac.uk/media/sites/researchwebsites/iridium/iridium_online_survey_report_17_8_2012_v2.1_SK.pdf

¹⁴ RDM policy analysis <http://iridiummrdrd.wordpress.com/2012/10/03/iridium-reporting-on-existing-internal-and-external-rdm-related-policy-analysis-mapping/>

The draft policy and code of practice brought before University Research Committee on the 10 December 2012. The Committee asked that further consultation with the Faculty Research Committees take place. Feedback received from this consultation related predominantly to the code of practice, which was subsequently amended.

3.2.4 Tools and systems

A broad discovery exercise and review of external RDM tools to meet local requirements was conducted. External tools were identified from the DCC website search, including the newly established 'Tools and Service' catalogue, those referenced in Jisc Managing Research Data Programme dissemination, the small number identified in the requirements gathering, and those known to project team members.

Tools were initially assessed for local institutional utility and integration with existing local technical infrastructure. A refined list of key tools going forward for evaluation was identified (see blog).

The range and capability of tools and systems to support RDM was found to be surprisingly immature and the availability of general solutions (such as technologies to support institutional repositories) was overestimated by the project team. This meant that consensus was not reached on pilot options until late in the project.

Technical effort was expended in the following directions.

- Development of a prototype research data catalogue. In broad terms this joins up data from two University research information management systems to associate projects with publications and to allow the addition of a small set of metadata and the location of data supporting the publication. This was undergoing additional user testing as the project drew to a close; a full specification is available from the [project outputs web page](#).
- Evaluation and customisation of the DCC's RDMP online system (see Appendix 8.5).
- Investigation of the use of the SWORD protocol to provide easy data deposit (see Appendix 8.3).

Following a conversation with Stuart Lewis, we attempted to implement the SWORD 'right click' desktop client¹⁵ but ran aground on Windows user interface issues.

We provided some assistance to the Bath Research360 project with the SWORD client they were developing for sakai, which is in use at Newcastle as a virtual research environment. Following on from that we looked at the possibilities of doing something with the SWORD libraries and sakai but concluded that such a development was almost a project in its own right.

- Investigation of the provision of e-science central¹⁶ (a cloud-based platform for data analysis) as a service, including the provision of a SWORD endpoint. The SWORD implementation is fully functional (create, read, update and delete of both metadata and files) and is described on the project website¹⁷; see Section 3.1 above for code locations. This implementation did raise an issue in that there is a problem with some of the operations resulting in the use of Java code provided by the SWORD standard website (swordapp.org) – it only became apparent during the system test phase that the code provided by them would not work for some situations.
- Identification of potential data repository solutions that could be used at the research group level if necessary.

Sharepoint was considered but we were unable to find any evidence of it currently being used to manage research data (rather than acting as a catalogue or research information management system). Given that the University has only a legacy Sharepoint service and that we would need to both manage expectations and produce a convincing business case for the expense of running up a full production service, we moved on.

¹⁵ <http://github.com/kshepherd/RightClickDeposit>

¹⁶ <http://www.esciencecentral.co.uk/> - a cloud-based platform for data analysis

¹⁷ http://research.ncl.ac.uk/media/sites/researchwebsites/iridium/iridium_e-Science_Central_SWORD_28_2_2013_v1_DM.pdf

Oxford's Dataflow¹⁸ systems were tried out quite early on. At the time we thought they were a bit too immature and didn't use a version of Linux that we support (see blog post¹⁹). We intend to revisit this in the future.

It wasn't until the October 2012 programme meeting and the emergence of CKAN as a possibility that we saw something that could possibly be developed to meet our needs. We investigated the API (see Appendix 8.4) and were able to set up a small pilot with a research group in Agriculture, Food and Rural Development, who are using it to archive their data (see [project outputs](#) for case study).

Very recently we have talked to vendors who are seeking to market curation solutions alongside their storage offerings, but remain unconvinced as to maturity and potential value for money; it is therefore likely that we will follow the open source route, at least for the time being.

3.2.5 Support

A range of activities and stakeholder mapping was conducted (see project plan). The user needs analysis from requirements gathering was fed into the human support infrastructure. Clearly there was a widespread requirement for updating and upskilling staff and research students in appropriate research data management approaches.

A review of existing training was undertaken with representatives of faculties, RES and Staff Development Unit (SDU). Locations for embedding RDM training in staff and student induction were identified. A website was developed in the institutional content management system Terminal 4 to support staff development in RDM (see Appendix).

Two 'writing days' were set aside for planning and writing documentation, with the following outputs:

- Re-drafted: policy principles and good practice guide
- Website wireframes plus some content (about, tools, etc.)
- Website development tools reviewed
- Tools guidance (How to: RDC) and other tools (outline/pointer)
- FAQs
- Good practice guide reviewed in detail
- Presentation dissemination
- Support implementation plan updated

An invitation to tender was issued and generated two responses. Netskills' tender was accepted and they were contracted to develop a 2-hour workshop with speaker notes and activities, and an on-line tutorial from information provided on the website. They were also tasked with interviews with students and staff to describe the importance of RDM techniques, and webcasts of common tools. Workshops were developed (see dissemination) to promote engagement with the outputs.

Human Factors Integration mapping identified the stakeholder 'who' and 'when', and their 'trigger' (what might cause them to seek information about RDM), with the 'tailoring' of workpackage outputs required. The trigger could be the submission of a funding proposal, the requirement to make metadata available to support a publication, staff development session or implementing good practice in a research team. The linking of stakeholders, with the project outputs in a timely fashion ensures a 'just in time' methodology. Thus outlining a plan for the transition of domain knowledge from the project and embedding of RDM good practice within the institution and its support post project.

3.2.6 Dissemination and engagement

Dissemination and engagement was conducted by regular updates including internal stakeholder engagement to the University Research Committee, University Research Forum, Faculty Research Committees through in person attendance and briefing papers.

¹⁸ <http://www.dataflow.ox.ac.uk/>

¹⁹ <http://iridiummrd.wordpress.com/2013/02/14/iridium-evaluation-of-datastage-and-databank-research-data-management-tools-from-dataflow-project/>

Wider University dissemination was carried out through the institutional 'NU Connect' newsletter and Registrar's consultation emails. Outputs and request for consultation were shared through the project website.

The project regularly engaged with the Jisc Programme through attendance and presentations at Programme Meetings, related DCC meetings and conferences. Highlights were:

- DCC Roadshow North East (presented)
- DCC Storage for RDM (attended)
- Digital Research Oxford 2012 (poster presented)
- Jisc Training Programme meeting (attended)
- ARMA Conference (poster presented)
- ECRM12 Conference (poster presented)
- KAPTUR Visual Arts and RDM (attended)

Specific engagement was carried out with the programme Research360 (University of Bath) project with a feedback email sent in March 2013 by Andrew Martin (ISS, Newcastle University) to Dr Catherine Pink and Jez Cope on the SAKAI platform technical work for RDM and community documentation.

The project blog posts were well received within the Programme and fostered sharing of outputs and a richer understand of the RDM domain²⁰.

The project has also been contact by the University of Massachusetts Medical School and Florida State University on RDM/librarianship and digital assets framework analysis respectively.

Blog posts	http://iridiummrdr.wordpress.com 53 blog posts written, 4211 page views and 42 comments as of 29 April 2013.
Project website	1915 unique page views ²¹ as of 29 April 2013.
Project Twitter feed	119 followers ²² , 320 tweets as of 29 April 2013.
Publicity materials	1 Newsletter, 14 draft posters, 2 'NU Connections' articles and 2 University Registrar consultation emails.

3.3 What did you learn?

- Online survey and interviews showed the diversity of research data types, locations, volumes, etc. and that this was not generally explicitly planned and managed in line with a formal research data management plan (RDMP) that would support best practice.
- Researchers were already involved in sharing research data (more often than had been anticipated).
- Finding individual research staff following the full research data lifecycle (from conception all the way through to long term retention and archiving if required) within a research project was not always immediately easy, with most focus and familiarity with data collection/analysis phase, however interviews helped identify those applying at least part of the pathway.
- A RDM policy was desired by the local research community to clarify expectations. The type of tools needed to support RDM was largely not expressed. Specific training in RDM had not been attended mostly and sources of training were not known to most.
- RDM tools available externally were either very specialised e.g. Datashield (Data aggregation through anonymous summary-statistics from harmonized individual-level databases) and Omero (specialised microscopy image repository) or too immature technically to implement as highly

²⁰ <http://iridiummrdr.wordpress.com/>

²¹ <http://research.ncl.ac.uk/iridium/>

²² https://twitter.com/iridium_mrd/

available production services. However, a lesson learned by the technical team is that high availability is not necessarily a major criterion for such tools. We also learned that, although there is nothing wrong with SWORD as a protocol, implementing it is difficult.

- Core metadata for research data sets was largely already available through existing research information management systems such as MyProjects and MyImpact and could be comparatively easily (and at low cost) combined with additional minimal metadata entry to populate a proof-of-concept Research Data Catalogue that could provide search functionality and for retrieving metadata.
- The postgraduate support team provided a pool of enthusiastic support with an understanding of specific research discipline approach and terminology. They were generally flexible in availability and could be directed to collectively complete tasks at comparatively short notice, thus providing a timely response to the project's immediate needs. Their location in Faculties provided guidance on approach and understanding of local environments.

There was an administrative burden in terms of recruitment, contracting and HR administration. The support team needed line management, final copy editing/review of tasks and varying degrees of coaching. During the course of the project some students required time away from duties for field work or for unanticipated leave due to demands of their studies, which required redistribution of project workloads.

- The need for policy is undermined without the appropriate infrastructure, however the policy is needed to justify the development of the infrastructure.
- Requirements and policy development were circular and iterative, in that for researchers to express what requirements they had, an outline of the policy principle expectations, roles and responsibility are needed to identify current gaps in practice. Requirements are required to populate the policy principles. Thus linear workpackages were not always beneficial. Lack of clear tools identification may have resulted from this.
- The term "research data management" has multiple connotations and is even unused in some disciplines (humanities and visual arts do not generally talk about "research data"). It is often confused with research data storage, assistance with which was often regarded as more important than assistance with the management and curation of stored data.
- The Staff Development Unit delivers general project management skills, research supervisory training and training in statistical analysis, but does not cover RDM and RDM planning. Additionally there was a requirement for database management.
- The Identification of 'carrots' to promote good practice in RDM was harder than 'sticks'. There remained a culture in some disciplines of researchers owning their data (that they often wanted to keep indefinitely) where the rewards and recognition for making data more discoverable and accessible needed to be clarified. Therefore achieving institution-wide change in practice to meet national standards required continued awareness of RDM through existing staff and student development programmes, website information and word of mouth dissemination. On-going consultation would also provide channels for embedding.
- A useful finding was the clarification and separation of active data and archival data (static) institutional requirements and implications.
- A Senior Management advocate at the highest level is required within the institution to drive wide-reaching implications of institutional RDM and is critical.
- REF2104 had a greater impact on the project than was anticipated.

3.4 Immediate Impact

- This project has raised awareness of the issues at all levels of the organisation.
- There is an aspirational policy available in draft which can provide a focus for future discussion and policy formulation.
- There is a dedicated support site for researchers with an email contact for enquiries

- Training of PI's, research associates, support staff and others involved in research is an important part of future RDM implementation across the institution. Iridium has supported the development of training materials that can be taken forward for adoption by SDU and the Faculties. Other materials can stand alone on, for example, the RDM webpages on the University website. Improved RDM planning by individual researchers can be supported by the DMPonline RDM planning tool and the authoring of a Newcastle-specific template
- Research Funding Development Managers are better equipped to deal with increasing demand for explicit RDM planning within funding applications.
- Project partners are more knowledgeable of the domain and confident in recommending what the institutional RDM approach should be.
- The immediate impact on the wider community was to contribute to the HE sector approach in responding to address institutional RDM through the Jisc Programme and to provide feedback and evaluation of RDM tools, support materials and strategy.

3.5 Future Impact

The project has raised the need for a strategy to implement RDM at the highest level. The future impact depends on institutional priorities for RDM in the context of the provision of other services in the institution. The current draft IT strategy refers to data curation and the Library's library strategic planning acknowledges the need for metadata services to support data curation.

- Development of the RDC functionality will ensure that data is discoverable in line with funder expectations and embedded in the day-to-day functions of the institution.
- RDM as a topic will have higher recognition
- Training materials in regular use and maintained/kept up to date by appropriate services
- One of the longer term outcomes of the project was the agreement by ISS to continue to develop CKAN as a potential basis for an institutional repository.
- The DMP Online system may also be implemented with a Newcastle hosted version.

4 Conclusions

Policy and future plans

- The project findings were able to underpin development of a business case for continuation, based on a conservative approach to forward planning, and aimed at maintaining the institutional national and international reputation and competitive edge, and avoiding future penalties.
- The draft institutional RDM policy and associated code of good practice for data management promotes compliance with current funder expectations (noting that the national policy landscape remained fluid), and makes the expectations, roles and responsibilities for the research community clearer by promoting alignment with best practice.
- The survey suggested that a majority of users' storage needs fell within the service currently provided by ISS; in order to promote good practice this could be provided as part of the overhead. For the remainder with significant needs, costs for data management and storage should be built into research proposals.
- Investment is required in RDM support across Central Services, Faculty and individual research groups.
- Increased research data discoverability may present unanticipated scenarios and governance opportunities.

Human factors

- Compliance with requirements for RDM, particularly DMP, has been greatly enhanced by the provision of information and signposting for researchers via the website.

- Support materials such as guidance on available external approved national data archives and discipline repositories should be tailored for integration and/or adaption by individual research group discipline needs.
- The content of the 2-hour workshop designed to be delivered to research students is relatively stable (not requiring extensive upkeep) due to being fairly generic within the current policy context.
- Embedding RDM in existing staff development programmes (one day research supervision and three day conducting research training) has given a pro-active dimension to human factors support.
- Staff recognise the potential benefits of access to datasets from elsewhere.
- Further embedding in the institutional infrastructure is desirable particularly via faculties with expert training given to research support staff on a 'training the trainer' model and introductions during staff and student induction.
- A formal requirement for RDM represents an opportunity to embed good practice in terms of data storage and curation.

Tools

- The CKAN data portal software shows promise as a data repository at the research group level, and the fact that instances can be federated could allow it to fulfil this function at the institutional level; we should therefore continue to develop it as a service. We should, however, also continue to investigate additional solutions as they arise – one size will not necessarily fit all.
- The Newcastle-specific DMP template encourages improved identification of research project responsibilities, pro-active addressing of research group and institutional technical and training infrastructures required and methods to maximise research impact. It provides a potentially valuable source of corporate information on current trends and on project needs and direction.
- A university-wide system with the functionality of the proof-of-concept RDC is required to support compliance with RCUK funder requirements on data discoverability (and pull together separate metadata i.e. from MyImpact and MyProjects).
- Simple end user tools to more accurately and easily cost individual RDM element resourcing (disk space, repository deposit, quality assurance, curation, retention duration) were needed to help research groups identify the implications of their needs in a timely manner, and these costs should be included into grant proposals where possible. Requirements should be identified during grant applications processes (using a light-touch approach) and advice sought if needs were outside of standard provision.

Metadata

After much discussion it was decided to promote simple metadata based on keyword and indexing of related publications (where possible) and free text entry. Social media approaches promoting frequently used words would be adopted, but formal taxonomies would not.

Licensing and ethics

Further work is needed in the area of confidential data and longitudinal studies, for example, in order to safeguard human participants and support funder-compliance with complex research studies.

Existing non-exclusive 'in perpetuity' licences were chosen to safeguard users and the owners of data set. Future access will require development of technical access management (access control).

5 Recommendations

The following recommendations are made with Newcastle University in mind but could be more generally applied.

- As outlined above, research data management consists of policy and practice, training and support, and technical support in respect of both IT functions and data curation. As these component functions are spread across a multiplicity of services, coordination is required; in

particular, joining up policy and practice with the technical support element is essential. In short, research data management needs an owner.

- The University needs to acquire central expertise in both digital curation and in long-term management of large quantities of data. In particular, the use of Digital Object Identifiers should be investigated.
- The University should ratify a research data management policy which will underpin both the expectations of funders and good practice in data management. Related to this is a need for policies regarding data retention and release together with a means of implementing them, and a requirement for suitable institutional licence(s) for released data.
- Guidance regarding the production of research data management plans should be provided for dissemination by the faculty Research Funding Development managers. In addition, data plan templates need to be generated and regularly reviewed. These could be supplemented by development of a version of the DCC's RDMP online system customised for the University.
- Support & training in RDM, based around good practice & planning, should be provided – particularly to postgraduate students and RAs. This could use the project outputs as a basis combined with a selection of the large volume of material that has been produced by the Jisc Research Data Management Training programme²³.
- In order to move towards funder compliance, the functionality provided by the prototype research data catalogue needs to be developed and enhanced by (at least) a search function and automatic metadata harvesting. This could be done by developing the current prototype, by developing the functions of the CKAN data portal, or by building it in to any redevelopment of the University's research information management systems.
- Support for tools used to manage workflows and active data during the course of a project should be provided. In particular, interfaces should be developed in order to facilitate transfer/deposit and this in turn requires a strategy for systems integration. Good practice in security and storage could be encouraged by the provision of an amount of storage to each research project.
- A centrally managed data repository, or set of repositories, should be considered for research data that needs to be retained locally after the conclusion of the project that generated it. Storage vendors are only now waking up to this requirement and so development of a supported open source solution such as CKAN seems to be the most likely scenario.
- The University should seek alliances and opportunities for shared services – for example: use of local and national Doctoral Training Centre network for delivery and coordination of RDM training to early stage researchers; shared services for data curation within the N8 Research Partnership²⁴.

As regards Jisc, we would recommend the following.

- Funder data management plan requirements and updates should be provided as a national service, including through data feeds/API provision for incorporation in local tools.
- DCC DMPonline tool to provide maximal, on demand, customisation and editing of an institutional template by the institution itself through access privileges.
- Emerging RDM tools should be monitored and reviewed, with recommendation made to academic community.
- Facilitation of a community of practice around the provision of research data management services; this would be particularly useful in areas such as CKAN development.

²³ http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/research-data-management-training.aspx

²⁴ <http://www.n8research.org.uk/>

6 Implications for the future

This project has provided a view of the state of the art of RDM and signposted possible ways forward. Its work should be of use to, for example, institutions which do not currently have a central data repository or a data catalogue.

Technical development work started as part of the project will be continued - for example an exploration of the scalability of CKAN to support research data curation and storage in a wider context than the pilot study referred to here. What is required is the establishment of a community with a common interest in this type of development.

It is currently not clear how the University intends to take RDM forward but most of the project outputs are in draft / prototype / pilot form and so do not predicate the final form of any RDM service built upon them. The support materials will be made available via the RDM web site (see Section 3.1) for the use of Faculty Research Funding Development Managers, the Staff Development Unit and any other interested parties.

7 References

See footnotes.

8 Appendices

8.1 Evaluation of external RDM tools and training

External tools assessment approach was blogged²⁵. The searchable DCC Tools & Service Catalogue (<http://www.dcc.ac.uk/resources/external/tools-services>) was released during the tools review and proved a useful service.

Table 1. Table of reviewed external RDM tools identified through DCC.

Tool	DCC Record	DCC Category	Target	Keywords	URL
Dataverse	http://www.dcc.ac.uk/search/node/Dataverse%20type%3Ares_tool	Managing Active Research Data	Researcher	Hosting, social sciences, versioning, access rights, harvesting, datasets	http://thedata.org/
Kepler	http://www.dcc.ac.uk/search/node/Kepler%20type%3Ares_tool	Managing Active Research Data	Researcher Curator	Scientific workflow, data curation	https://kepler-project.org/
LabTrove	http://www.dcc.ac.uk/search/node/LabTrove%20type%3Ares_tool	Managing Active Research Data Workflow and Lab Notebook Management	Researcher	Research blogging, electronic notebook, instrumentation, metadata capture, annotation, laboratory	http://www.labtrove.org/
MyExperiment	http://www.dcc.ac.uk/search/node/MyExperiment%20type%3Ares_tool	Managing Active Research Data	Researcher	Research social networking, collaboration, sharing, workflows, plans, digital objects	http://www.myexperiment.org/
Taverna	http://www.dcc.ac.uk/search/node/Taverna%20type%3Ares_tool	Managing Active Research Data	Researcher	Large-scale data, services, scripts, computational tasks, modeling, workflow management	http://www.taverna.org.uk/
WebCite	http://www.dcc.ac.uk/search/node/WebCite%20type%3Ares_tool	Managing Active Research Data Creating unique identifiers	Researcher Publishers	Web archiving, snapshot, digital objects, unique identifier	http://www.webcitation.org/
Researchgate	http://www.dcc.ac.uk/search/node/Researchgate%20type%3Ares_tool	Sharing Output and	Researcher	Networking, outputs, profile	http://www.researchgate.net/

²⁵ <https://iridiummrd.wordpress.com/2012/10/03/iridium-reporting-on-identification-of-available-external-and-internal-tools-rdm-tools/>

Tool	DCC Record	DCC Category	Target	Keywords	URL
	1	Tracking Impact			
Curator workbench	http://www.dcc.ac.uk/search/node/Curator%20workbench%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Data preparation, pre-submission, repository submission	http://www.lib.unc.edu/blogs/cdr/index.php/2011/01/13/curators-workbench-is-now-free-and-open-source-software/
Archivelt	http://www.dcc.ac.uk/search/node/Archivelt%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Archiving, records, snapshot	http://www.archive-it.org/
Duke Data Accessioner	http://www.dcc.ac.uk/search/node/Duke%20Data%20Accessioner%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Data migration, physical media, server, checksum, migration workflow	http://library.duke.edu/uarchives/about/tools/data-accessioner.html
FITS	http://www.dcc.ac.uk/search/node/FITS%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Metadata, curation, validate, extract, normalising	http://code.google.com/p/fits/
Heritrix	http://www.dcc.ac.uk/search/node/Heritrix%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Web archiving, web crawler	https://webarchive.jira.com/wiki/display/Heritrix/Heritrix
JHove2	http://www.dcc.ac.uk/search/node/JHove2%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Digital object characterise, format, validating, metadata extraction	https://bitbucket.org/jhove2/main/wiki/Home
MIXED	http://www.dcc.ac.uk/search/node/MIXED%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Transformation, XML, spreadsheets, database, SDFP	https://sites.google.com/a/datanetworkservice.nl/mixed/
Nesstar	http://www.dcc.ac.uk/search/node/Nesstar%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Data publication, platform	http://www.nesstar.com/
Netarchive Suite	http://www.dcc.ac.uk/search/node/Netarchive%20Suite%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Web archiving, harvesting, consistency check	https://sbforge.org/display/NAS/NetarchiveSuite;jsessionid=13C5C9FAB386702BF776522C3E345DEE
NLNZ Metadata Extraction Tool	http://www.dcc.ac.uk/search/node/NLNZ%20Metadata%20Extraction%20Tool%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Metadata extraction, digital file, XML	http://www.natlib.govt.nz/services/get-advice/digital-libraries/metadata-extraction-tool/?searchterm=metadata%20extraction
PREMIS in METS Toolbox	http://www.dcc.ac.uk/search/node/PREMIS%20in%20METS%20Toolbox%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	PRESMIS, METS, validation, descriptor	http://pim.fcla.edu/
The BagIt Library	http://www.dcc.ac.uk/search/node/	Depositing and	Curator	Packaging, digital	http://sourceforge.net/projects/loc-xferutils/

Tool	DCC Record	DCC Category	Target	Keywords	URL
	The%20BagIt%20Library%20Tool%20type%3Ares_tool	Ingesting Digital Objects		container,	
Web Curator Tool	http://www.dcc.ac.uk/search/node/Web%20Curator%20Tool%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Web archiving, ARC format	http://webcurator.sourceforge.net/
Xena Software	http://www.dcc.ac.uk/search/node/Xena%20Software%20type%3Ares_tool	Depositing and Ingesting Digital Objects	Curator	Digital preservation, transformation, conversion, open format	http://xena.sourceforge.net/
ICA-Atom	http://www.dcc.ac.uk/search/node/ICA-Atom%20type%3Ares_tool	Archiving and Preserving Information Packages	Curator	Standard description, archival holdings, publishing	http://ica-atom.org/
XArch	http://www.dcc.ac.uk/search/node/XArch%20type%3Ares_tool	Archiving and Preserving Information Packages	Curator	Merging, multiple versions, nested merge	http://xarch.sourceforge.net/
CONTENTdm	http://www.dcc.ac.uk/search/node/CONTENTdm%20type%3Ares_tool	Archiving and Preserving Information Packages	Curator	Web server, discovery	http://www.contentdm.org/
Curators Workbench	http://www.dcc.ac.uk/search/node/Curators%20Workbench%20type%3Ares_tool	Archiving and Preserving Information Packages	Curator	Archiving, deposit, metadata organisation	https://github.com/UNC-Libraries/Curators-Workbench
D-Net	http://www.dcc.ac.uk/search/node/D-Net%20type%3Ares_tool	Archiving and Preserving Information Packages	Curator	Repository networking	http://www.d-net.research-infrastructures.eu/
Datecite	http://www.dcc.ac.uk/search/node/Datecite%20type%3Ares_tool	Archiving and Preserving Information Package	Curator	Data citation	http://datecite.org/
Dioscuri	http://www.dcc.ac.uk/search/node/Dioscuri%20type%3Ares_tool	Archiving and Preserving Information Package	Curator	Preservation, hardware emulator	http://dioscuri.sourceforge.net/dioscuri.html
Dspace	http://www.dcc.ac.uk/search/node/Dspace%20type%3Ares_tool	Archiving and Preserving Information Package	Curator	Repository system	http://dspace.org

Tool	DCC Record	DCC Category	Target	Keywords	URL
DuracLOUDs	http://www.dcc.ac.uk/search/node/DuracLOUDs%20type%3Ares_tool		Curator	Cloud service, interface	http://www.duracLOUD.org/
Eprints	http://www.dcc.ac.uk/search/node/Eprints%20type%3Ares_tool		Curator	Repository system	http://www.eprints.org/
Fedora	http://www.dcc.ac.uk/search/node/Fedora%20type%3Ares_tool		Curator	Repository system	http://fedora-commons.org/
JHOVE	http://www.dcc.ac.uk/search/node/JHOVE%20type%3Ares_tool		Curator	Validation, digital object, format	http://hul.harvard.edu/jhove/
KEEP Emulation Framework	http://www.dcc.ac.uk/search/node/KEEP%20Emulation%20Framework%20type%3Ares_tool		Curator	Preservation, Software emulation,	http://emulframework.sourceforge.net/about.html
LOCKSS	http://www.dcc.ac.uk/search/node/LOCKSS%20type%3Ares_tool		Curator	Subscription content, digital collections	http://lockss.org
MIXED	http://www.dcc.ac.uk/search/node/MIXED%20type%3Ares_tool		Curator	Data transformation, tabular data	https://sites.google.com/a/datanetworkservice.nl/mixed/
Creative Commons tools	http://www.dcc.ac.uk/search/node/Creative%20Commonstools%20type%3Ares_tool	Managing and Administering Repositories	Curator	Copyright, licensing	http://creativecommons.org/
OpenDOAR	http://www.dcc.ac.uk/search/node/OpenDOAR%20type%3Ares_tool	Managing and Administering Repositories	Curator	Repositories catalogue	http://www.opendoar.org/index.html
Tufts Submission-Agreement Builder Tool	http://www.dcc.ac.uk/search/node/Tufts%20Submission-Agreement%20Builder%20Tool%20type%3Ares_tool	Managing and Administering Repositories	Curator	Submission agreement creation	http://sites.tufts.edu/dca/about-us/research-initiatives/taper-tufts-accessioning-program-for-electronic-records/

Additionally RDM tools were identified locally and through Jisc MRD02 Programme dissemination and were further categorised.

Table 2. Table of project and Programme identified RDM tools.

Tool	Category	Target	Summary	Keyword	URL
DMPOne v3	Planning	Researcher	Online template system for data management planning.	Funders, policy, planning, data	https://dmponline.dcc.ac.uk/

Tool	Category	Target	Summary	Keyword	URL
				management plan, template	
CARDIO	Audit/survey	Service Managers	RDM infrastructure assessment	Self-assessment, infrastructure maturity	http://cardio.dcc.ac.uk/
EduServ cloud	Storage	Researcher	Cloud storage infrastructure as a service	Cloud	http://www.eduserv.org.uk/
Datacite	Archiving/citation	Researcher	Assigns persistent identifiers to datasets to allow greater ease of citing datasets as sources in publications.	Citation, DOI, persistent citation	http://datacite.org/
DAF (Digital asset framework)	Audit/survey	Service Managers	Research data assets assessment/survey.	Self-assessment, audit	http://www.dcc.ac.uk/resources/repository-audit-and-assessment/data-asset-framework
Simple Web-service Offering Repository Deposit (SWORD)	Standard/protocol	Repository/System Managers	Research data/metadata transfer tool	Data transfer	http://swordapp.org/about/
ViDass	Online cloud service and RDM	Researchers	Facilitation of research database re-use.	Relational database	http://vidaas.oucs.ox.ac.uk/
BRISKit (see below also)	Clinical data hosting service	Researcher	BRISKit will develop a national shared service brokered by JANET to host, implement and deploy biomedical research database applications that support the management and integration of tissue samples with clinical data and electronic patient records	Tissue samples, clinical studies, NHS	http://www2.le.ac.uk/offices/itservices/resources/cs/ps/o/project-websites/brisikit
CiviCRM	<i>Customer relationship management</i>	Researcher	Web-based, internationalised suite of computer software for constituency relationship management.	Clinical data, NHS, record keeping	http://civicrm.org/
REDCap	Build and	Researcher	Secure web application (REDCap)	Clinical data, NHS,	http://project-redcap.org/

Tool	Category	Target	Summary	Keyword	URL
	manage data bases		designed exclusively to support data capture for research studies	record keeping	
OBiBa Onyx	Data storage management and Application server Intranet	Researcher	Onyx stores the data collected during the different stages centrally and makes it available to all workstations.	Clinical data, NHS, record keeping	http://www.obiba.org/node/3
i2b2	Research data warehouse	Researcher		Clinical data, NHS	https://www.i2b2.org/
KRDS (Keeping research data safe)/ Beagrie value tool chain	Cost-benefit analysis	Curator Researcher	Provides a platform from which organisations can identify, analyse and communicate the benefits of investing in research data management and storage. The Beagrie Value Chain represents a more sophisticated version of the KRDS Benefits Framework, and the two can be used together.	Benefits, evaluation	http://www.beagrie.com/krds.php
Metadata extractor	Preservation of metadata	Curator	Programmatically extract preservation metadata from a range of file formats like PDF documents, image files, sound files Microsoft office documents, and many others	Metadata	http://meta-extractor.sourceforge.net/
Zenity	Managing and Administering Repositories	Curator	Zenity is a research output repository platform that provides a suite of building blocks, tools, and services that help to create and maintain your organization's digital library ecosystem.	Research outputs	http://research.microsoft.com/en-us/projects/zenity/
DataStage/ DataFlow	Local file management	Researchers	Secure personalized 'local' file management environment for use at the research group level, appearing as a mapped drive on the end-user's computer	File management, data collaboration, data deposit	http://www.dataflow.ox.ac.uk/index.php/about/about-datasetage

Tool	Category	Target	Summary	Keyword	URL
MyLabBook	Active data	Researcher	Open source online laboratory notebook	Digital lab book, documentation	http://mylabbook.org/
CARMEN	Data curation	Researchers	CARMEN is a web based data management project based on SRB (Storage Resource Broker)	Neuroscience. Data access, data curation	http://www.dcc.ac.uk/resources/case-studies/carmen-0
rsnapshot	Backup	Researcher	File system snapshot utility for making backups of local	Linux, snapshot, backup	http://rsnapshot.org/
CasaXPS	Powerful analysis techniques for both spectral and imaging data.	Researcher	CasaXPS software for conversion to the ISO standard format results to be easily exchanged.	Data transformation	http://www.casaxps.com/
ArcGIS	Cloud server system for the data.	Researcher	A complete system for designing and managing solutions through the application of geographic knowledge.	Geo data	http://www.esri.com/software/arcgis/features.html
Dotmatics Browser and Gateway	Powerful querying and reporting tool.	Researcher	Integrates data from any database whether it is chemical, biological, technical etc. Create and share lists, queries and forms across projects or research teams. Combines with other Dotmatics' solutions to provide a complete and seamless data management and visualisation system.	Data management, visualisation	http://www.dotmatics.com/products/browser/
Dropbox	Managing Active Research Data	Researchers	Main function in relation to RDM is storage. Any file saved to Dropbox will automatically be saved to all of that user's devices – this provides multiple copies of the data and prevents problems with versioning. Also gives user the ability to share data easily with	Data sharing, data access	https://www.dropbox.com/

Tool	Category	Target	Summary	Keyword	URL
			other team members so that data does not need to be stored in emails or on USB drives.		
SparkleShare	Managing Active Research Data	Researcher	'Dropbox' with Git functions. Discussed at Jisc MRD hack day.	Data sharing	http://sparkleshare.org/
<i>Recent tools noted post original review</i>					
ZendTo	Active Data	Researcher	Secure file transfer.	Security, large files, data sharing	http://zend.to/
ISA tools	Metadata	Researcher	Metadata/scientific context tool.	Scientific context, metadata, data re-use	http://www.isa-tools.org/
CKAN	Data publishing/hosting	Researcher	Data publishing/repository tool.	Repository	http://ckan.org/
DMP20	Planning	Researcher	Simple data management planning tool.	Data management planning	http://www.miidi.org:8040/dmp/

8.2 Research Data Catalogue user testing

The background to the proof-of-concept Research Data Catalogue (RDC) has previously been described²⁶.

Staff members from a specific institutional research theme were invited to use the system within a 4 week period, with sample research information data (projects, publications) from their own research profile imported to the system. Testers were asked to link publications to the appropriate research project (where possible) and identify where the research data was.

Representative quotes from testers:

I have had a go at using the Research Data Catalogue and on the whole it seems quite straightforward to use.

I found the website easy to use, I did not encounter any problems.

I did try and link it to the data sources with varying degrees of success and failure. I have a few comments/feedback which may be useful.

Table 3. Themes and implications from comments received from user testers.

User feedback themes	Implications
Work was not funded through a grant	Representation of 'unfunded' projects/publications within University system should be considered
Webpage needs a help section to explain what is required for each field	System (both website and offline) documentation should give further guidance and examples of good practice in describing records
Function and definition of each 'button'/feature	Documentation should explain better system functions (i.e. 'filter', 'metadata status')
Publication data located in multiple locations	Practical guidance need on how to record publication data sets. 'Packaging' of data sets and archiving in a set location should be considered
Single data location record and other information needs to be re-used for multiple publications	Common user responses should be selectable from a template/common settings profile to save time
Efficient entry of the metadata	Use of quality metadata source to pre-populate records should be considered (i.e. journal publication keywords)
'Auto-tagging' metadata quality	System suggested 'auto-tagging' needs to be more sophisticated
User reassurance on record entry completion	Usability to be improved by making save record feature give more overt confirmation
Greater filtering of different publication types	More advanced filtering needed where possible from underlying database records
User role delegation (i.e. Principal Investigator to Co-investigator)	Role delegation functions from similar RIM systems should be considered
Imported database records up to date	Needs to be investigated if specific to testing process or reliant on external data feed provider/a reliant system publication claim process

The RDC final specification is available from project outputs page²⁷.

²⁶http://research.ncl.ac.uk/media/sites/researchwebsites/researchdatamanagement/iridium_Research%20Data%20Catalogue%20Quick%20Start%20Guide_April_2013_V6_user-testing.pdf (accessed June 2013)

²⁷http://research.ncl.ac.uk/media/sites/researchwebsites/iridium/iridium_research_data_catalogue_specification_07_6_2013_v1_PT.pdf (accessed June 2013)

8.3 SWORD evaluation

The work carried out in relation to SWORD specification was essentially a completely open investigation of the SWORD specification and its potential fit and ease of implementation for use with research data management and related repositories and technologies.

In order to fully understand and assess the SWORD protocol you must first understand the conglomerate of technologies that make up the spec. In essence SWORD is a “profile” of the ATOM publishing protocol specification that also uses the Internationalised Resource Identifier (IRI) specification and Basic Auth. SWORD also has 2 variants, the original v1.3 and the later v2 specification, this investigation mainly focussed on version 2 but did not ignore for the older spec for reasons of understanding the specifications evolution. In addition the SWORD website provided a lot of useful introductory materials.

The majority of work, after assessing the various technologies, revolved around trying to understand the example implementations and libraries written in the java programming language, this instance was chosen fairly arbitrarily as a personal preference as one of the advantages of the sword profile is its relative agnosticism to underlying technologies. The libraries come in two flavours, server and client, i.e one allowing a repository to talk SWORD and the other talking SWORD to a repository, respectively.

Getting first principles to work was surprisingly easy, but it quickly became clear that the ease of implementation from thereon in was reliant on two things: 1) Your overall understanding of the parameters passed and their usages in repositories and 2) if your chosen language has an element of http file transfer that is easy, uncomplicated, well documented and correctly implemented. These two points, it would seem, are the true dependencies to success in implementing SWORD interactions and therefore it would seem wise to also invest time in becoming familiar with additional external concepts such as Dublin core and the intricacies of file transfer for your specific choice of language.

Just as the client/server work was coming to the above two conclusions a blog post was contributed back to the community as feedback of experiences so far
(<http://iridiummrd.wordpress.com/2012/10/04/sword-v2-from-clueless-to-claymore/>)

The overall impression gleaned from SWORD was a well thought through protocol that’s only main lacking area is that of good, complete, examples of how to apply the technology best, but this is probably indicative of any relatively immature protocol. It should be noted that the profile assumes you are coming from a background familiar with repositories and related technologies, starting with SWORD and working back into the repositories is quite an uphill challenge!

As a final recommendation, I have to wonder if either/or:

- The protocol could be abstracted a further level away from the atom publishing protocol and just purely concentrate on how to standardise the packaging of depositing (over and above concepts such as Dublin core) and let the implementer choose the transport medium (i.e. SOAP or REST), this may also be more conducive to integrating authentication mechanisms other than basic auth.
- It might be worth considering a “lite/basic” version of SWORD that purely does a very common type of deposit, for example unauthenticated deposit to a predefined place in a repository, even if just for exemplary purposes.

Andrew Martin
June 13

8.4 CKAN API

The work in relation to the CKAN API was a similarly open ended investigation to the SWORD work but since the realm of integrations and CKAN (as it turned out) seems even less mature than SWORD, it ended a little more fruitfully with actual software contribution back to the community.

This package of work started with an assessment of what the API was capable of (and by extension what CKAN itself was capable of), the API is split into several versions and programmatic approaches. Versions 1 and 2 employ a REST-ish approach and the later version 3 adopts a style closer to SOAP than REST, which may have HTTP verbs and JSON, but splits functionality up into

URLs that do very specific things for specific objects. All versions, as you would expect, follow the idea of passing objects for internal CKAN concepts such as users, groups, datasets etc...

Upon looking for clients to the API's it was found that there is quite a reasonable spread of prewritten clients in various languages, I concentrated most of my effort on looking at the PHP and Java implementations; however, it became quickly obvious that the client code base had fallen somewhat behind the server code base and the clients needed updating to the version 3 approach. Thankfully a stub of a client already existed and from that I began to extend and reconfigure the code with the intent of forming reusable patterns (i.e. standardising a series of gets and sets to concepts and then trying to incorporate "Object factories") that could be easily ported and similarly update the PHP client. This source was contributed to the community via github <https://github.com/andmar8/CKAN-Java-Client>.

As an extension of this work, some very preliminary work was carried out to create a sakai tool to try and test bed the client code. The sakai development environment, however, is not a small project to figure out in itself, so I biased the majority of time on furthering the CKAN client.

Overall impressions of working with the CKAN API was it is refreshingly sane and given the pseudo-REST approach is relatively simple to interact with, possibly mirroring the python-esque culture in the CKAN background. This does have its downsides (or possibly a "community naivety") though in that there seems to be an overall approach of ignoring the difficulties of interacting with a potentially highly dynamic API from very popular but lesser dynamic/non-prototypal languages.

In conclusion, from a developer's perspective, CKAN is well documented and seems relatively stable, but a (reportedly) fluid API structure could cause problems on going for integrators of strongly typed and/or lesser dynamic languages (JAVA/C#/PHP), however my understanding of the "more fluid" aspects of the API are geared more toward user extensions of the API, in which case you would hope persons extending the API would be careful enough to document and write clients that support those extensions. Initial attempts at community support was also slightly troubling as the time taken to respond can be somewhat lengthy, but I think that is probably more a facet of the small size and workloads of the development community than anything else.

Andrew Martin
June 13

8.5 DCC DMPonline (v3) tool evaluation

8.5.1 Background

The web-based DMPonline (v3)²⁸ research data management (RDM) tool, developed by the Digital Curation Centre (DCC), was progressively evaluated between November 2011 and June 2013 to support the draft RDM policy and local good practice. It should be noted that the tool is in continual development and an updated version of the tool (v4) is expected in August 2013.

DMPonline supports writing research data management plans (RDMP) for most RCUK funders and 2 major biomedical charitable funders, where template guidance is provided by the funder. Its use is formally recommended by certain funders such as the MRC.

8.5.2 Findings summary

DMPonline is a mature and comparatively easy tool to use. It proved a useful in supporting data management planning. Provision of available RCUK funders' templates and, importantly, in context further good practice information and DCC Checklist guidance was valuable. The templates seem up to date with recent changes to the NERC plan requirements implemented promptly.

By the nature of an online system it can support the sharing/collaborative aspect of RDMP. The tool allows export of users RDMPs in common file formats for offline use and further editing of data management plans which is still preferable for many users. Moreover, some formatting of exported plans will likely be required to work within funder page length limits and readability.

²⁸ <https://dmponline.dcc.ac.uk/>

Development of a local institutional template (post-award) was required as the available RCUK directed (or inferred) templates only account for approximately one third of research projects funded at the institution (for example the EPSRC is a notable exception without a mandated template). An institutional template was authored after reviewing other RCUK templates across research disciplines for the most pertinent question themes with additions to support local institutional priorities (e.g. identification of RDM resourcing, not duplicating existing workflows on ethics permissions, etc.).

Local experience of using the DMPOnline tool features was written up as a local user manual²⁹.

8.5.3 DMPOnline recommendations

We recommend that addressing of the follow issues will aid DMPOnline widespread utility both locally and nationally.

General recommendations

- Noting existing documentation updated during evaluation period (DMPOnline Quick Guide³⁰, User Guide³¹) and older screencast³², a full step-wise user manual is preferred from experience locally to give confidence that end users were aware of all system functionality and correct use.
- End users need advance notification of any changes to tool function, features and funder templates to allow for local resource planning and dissemination. A user subscribers' email list to announce these changes is needed as a priority.
- Particularly for users familiar with using official funder website directly downloaded template versions, reproduction within the DMPOnline tool and exported plan template needs to be nearly identical for user acceptance.
- When a plan is shared with other users, notifications and history/versioning of changes made by different users should be stronger.
- Exportation of guidance notes, in context of template headings, from tool 'information' pop-up windows would be useful in the exported template as this seems to be a common end user preferred way of working.
- When completing a plan, you can see the progress bar, but the progress bar does not tell you which sections need completing or which questions were not answered, thus you can easily get lost. Therefore greater tracking of progression is needed.
- Specific usability enhancements such as section heading titles navigation, or 'hover over' information, in addition to section numbers for easier selection of sections, and clarifying the locking/duplication workflows, etc. These will be proposed on DCC DMPOnline Github feedback page³³.

Specific recommendations on implementing an institutional RDMP

- A greater level of tool administrator privileges for an institutional template moderator in DMPOnline are needed so institutions can edit, update and revise templates on demand themselves. This immediacy in making changes is essential.
- A DMPOnline API for integration of up to date template content within institutional local systems, in collaboration with Funders, as a supported national service, is desirable.
- Creation/uploading of novel templates by the end user would be a desirable and popular feature.
- More direct access and channelling to the institutional specific template via use of a specific URL or linked to login details is required.

²⁹

http://research.ncl.ac.uk/media/sites/researchwebsites/iridium/iridium_NCL_DMPOnline_guidance_DRAFT_v6_JW_LW.docx

³⁰ https://dmponline.dcc.ac.uk/system/attachments/12/original/DMP_Online_Quick_Guide.pdf?1350652998

³¹ https://dmponline.dcc.ac.uk/system/attachments/14/original/DMP_Online_User_Guide.pdf?1350652948

³² <http://www.screenr.com/Syo>

³³ <https://github.com/DigitalCurationCentre/DMPOnline/issues>

- Streamlining of numerous information ‘i’ buttons from DCC Checklists, external and institutional guidance in several places of the onscreen display, within a template section, can sometimes become confusing.
- Question ‘logic’ of linked, follow on questions, needs to be more robust to remove questions not relevant based on previous responses

8.5.4 Wider institutional data management planning implications

The increasing requirement for data management plans and an anticipated strengthening of the review process leads to some notable points for consideration.

- Formally documenting ‘end-to-end’ RDMP is new to many researchers and support will be needed.
- Robust data management planning takes time, particular if the process had not been documented previously. A standard RDMP for each research group, to be adapted for specific individual research project proposals would be beneficial. Sharing of RDMPs, with their examples of good practice, and local peer review would be beneficial.
- Elements of RDMP such as resourcing costs need to be considered at the earliest stage possible. Ethical considerations already appear as a ‘flagged’ question early in institutional research information management pre-award online workflow processes (i.e. MyProjectsProposals). Key elements of RDMP need to be similarly addressed.
- Arguably the current *iridium* project draft RDMP template if followed likely leads to more robust institutional RDM planning as it has strong focus on resourcing, collaboration and auditing – exceeding that required in some RCUK provided templates.
- Estimating the RDMP needs based on number of grant applications and successfully awarded projects, across various Funders and Faculties, will aid institutional resource planning.
- Information contained within RDMP plans has the potential to be used for better future institutional infrastructure and resource planning if it can be collated and analysed.