UNIVERSITY OF BATH

Citation for published version:
Ball, A 2012, *Tools for research data management*. University of Bath, Bath, UK.

Publication date:
2012

Link to publication
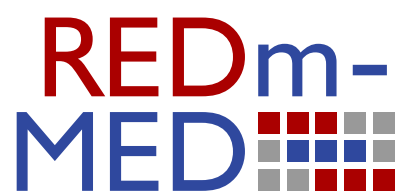
University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Tools for Research Data Management

Alex Ball

redm2rep120110ab10.pdf

issue date: 16th March 2012

Catalogue Entry

Title	Tools for Research Data Management
Creator	Alex Ball (author)
Subject	research data management; data repositories; training materials
Description	The REDm-MED Project seeks to provide recommendations to researchers within the University of Bath's Department of Mechanical Engineering on tools and resources they could use to ensure compliance with the Department's Research Data Management Plan Requirements Specification. In order to inform such recommendations, this document reviews some of the techniques, tools and training resources produced by projects in the JISC Managing Research Data Programme.
Publisher	University of Bath
Date	10th January 2012 (creation)
Version	1.0
Type	Text
Format	Portable Document Format version 1.5
Resource Identifier	redm2rep120110ab10
Language	English
Rights	© 2012 University of Bath

Citation Guidelines

Alex Ball. (2012). *Tools for Research Data Management* (version 1.0). REDm-MED Project Document redm2rep120110ab10. Bath, UK: University of Bath.

1 INTRODUCTION

The REDm-MED Project seeks to provide recommendations to researchers within the University of Bath's Department of Mechanical Engineering on tools and resources they could use to ensure compliance with the Research Data Management Plan Requirements Specification [DB12]. In order to inform such recommendations, this document reviews some of the techniques, tools and resources produced by projects in the JISC Managing Research Data Programme. Specifically the projects in the Research Data Management Infrastructure (RDMI), Research Data Management Planning (RDMP), Citing, Linking, Integrating and Publishing (CLIP) and Research Data Management Training Materials (RDMTrain) strands are considered. The tools related to the Support and Tools strand – CARDIO and the KRDS Benefits Analysis Toolkit – are already well known and are not covered here; similarly, the outputs of the ERIM Project are not reviewed as they already form the basis for the REDm-MED Project.

2 RESEARCH DATA MANAGEMENT INFRASTRUCTURE PROJECTS

2.1 ADMIRAL and DataFlow

The ADMIRAL Project was lead by the University of Oxford.[1] It aimed to provide both an environment for organizing, collaborating on and annotating biological datasets, and tools to assist in depositing annotated datasets in the Oxford University DataBank, the institution's data repository. The tools developed by ADMIRAL have been taken forward by the JISC DataFlow Project with funding from the Universities Modernization Fund.[2]

DataStage. DataStage is a file storage environment for working with research data. It is designed to be deployed at the level of a research group. It appears like a normal network drive on users' computers, but with additional controls for configuring access to directories on the drive, allowing for private, shared, collaborative, public and communal areas. As well as this mapped-drive approach (using the Common Internet File System protocol, or CIFS), the file store can also be accessed via SSH/SCP, SFTP, WebDAV, and through a Web interface. This latter interface contains a data packaging tool for converting a directory (and all its subdirectories) into a BagIt-compliant ZIP file and submitting it to a data repository using the SWORD v2 protocol [Jon; Wil12]. DataStage is implemented as a virtual Linux-based server, although it can delegate the actual storage of files to cloud services like DropBox. It is configured to run regular automatic backups, and uses LDAP (synchronized with a couple of other mechanisms) for authentication.

DataBank. DataBank is a redeployable version of Oxford's institutional data repository. It is written in Python and runs on a virtual Linux server, and thus can be installed locally or as a cloud service. DataBank is designed to run several data silos in parallel; each can be administered separately and given a specialized search interface. Among the services it provides are ingest via the SWORD v2 protocol, DOI registration via DataCite,[3] version control (with symbolic links between unchanged files, for storage efficiency), embargo

---

1. ADMIRAL Project Website, URL: http://imageweb.zoo.ox.ac.uk/wiki/index.php/ADMIRAL
2. DataFlow Project Website, URL: http://www.dataflow.ox.ac.uk/
3. DataCite Consortium Website, URL: http://datacite.org/

management, metadata exposure via OAI-PMH [OAI08a], and integration with file format services such as JHOVE and DROID.[4,5]

2.2 FISHnet

The FISHnet Project was a collaboration between the Centre for e-Research at King's College London and the Freshwater Biological Association with the aim of setting up a data repository and rights framework at the latter organization.[6] While the project did not produce any tools that might be reused by REDm-MED, there are a few aspects of its work that may prove instructive.

The project employed a traffic-light system to represent the different levels of curatorial attention being paid to the datasets ingested into the repository [Haf10a].

Red category. At a minimum, authors upload data (or provide a link to where it resides in another repository) in an arbitrary state, that is, not necessarily in a reusable form. They also provide basic metadata and their contact details. The data are indexed and users of the repository can search for them, though they must contact the depositor for more information or for a copy of the data.

Amber category. To achieve this category, the data must be uploaded in a reusable form, and the depositor must provide more extensive metadata. The 'reward' for this extra effort is that the data are given a DOI to enable them to be cited, though by default users must register with the system before being given direct access to the data. The repository also commits to preserving the data to maintain its usability.

Green category. To move from the Amber category to Green, data must be released under a highly permissive license or waiver, such as Creative Common Zero [CC]. Once this is done, all access restrictions are removed. Data in this category may be converted to RDF and entered into a triple store to allow queries via SPARQL and integration with the rest of the Semantic Web [W3C04; W3C08].

The Freshwater*Life* repository was built using Fedora Commons as the digital repository backend,[7] Liferay as the user-facing frontend,[8] with access control handled by OpenAM[9] and search by Apache Solr[10] [Haf10c]. The Project used the Scrum method of agile development, with issues tracked in Jira[11] [Haf10b].

2.3 I2S2

The I2S2 Project was a collaboration between the universities of Bath, Southampton and Cambridge, and the Science and Technologies Facilities Council (STFC). The aim was to improve the level of integration between institutions and national-level services in order to streamline researcher workflows in fields such as Chemistry and Earth Sciences.

---

4. JHOVE Website, URL: http://hul.harvard.edu/jhove/
5. DROID development Website, URL: http://droid.sourceforge.net/
6. FISHnet Project Website, URL: http://www.fishnetonline.org/home
7. Fedora Commons Website, URL: http://fedora-commons.org/
8. Liferay Portal Website, URL: http://www.liferay.com/
9. OpenAM Web page, URL: http://www.forgerock.com/openam.html
10. Apache Solr Web page, URL: http://lucene.apache.org/solr/
11. JIRA product Web page, URL: http://www.atlassian.com/software/jira/

Aside from cost–benefit analyses and a sustainability plan, the three major outputs of the project were the *I2S2 Idealised Sceintific Research Activity Lifecycle Model* [I2S11], an improved administration system for the National Crystallographic Service, and ICAT-Personal,[12] a version of STFC's ICAT data catalogue suitable for installation and use within a research group [YM11].

The particular use case ICAT-Personal seeks to serve is to capturing a processing workflow from raw data to publishable data. It uses a Java database for storage. Ingest is via an XML wrapper (using a custom schema), interpreted by a tool called XMLExtractor; the wrapper indicates which programs were used in the workflow, and in what order, and for each program, the data inputs, the configuration parameters, the output data and auxiliary outputs. Once ingested, the workflow may be visualized as a Graphviz dot file (or derivative graphic) using the DotGeneration tool.[13] Finally, the DataGrabber tool is provided for retrieving ('restoring') data from the catalogue.

2.4 IDMB

The IDMB Project was lead by the University of Southampton, with partners including the Naitonal Oceanography Centre and the University of Oxford.[14] The major outputs of the project were a three-level metadata structure for research data, improved storage capacity, a ten-year roadmap for developing the data management infrastructure, and an institutional data management policy and business case [BPW11].

The three levels of metadata are as follows [Tak10].

1. *Core.* These are metadata optimized for discovery. The four key fields are Author, Discipline, Publisher, Date. The IDMB Project decided to use Dublin Core metadata at this level [DCM10].

2. *Discipline.* These are metadata useful for those in the given discipline for classifying a collection of resources. Fields may include sub-domains, projects, funders, or techniques, but will typically be discipline-specific.

3. *Project.* These are detailed metadata for individual resources within a collection. The fields that are relevant, and the way of expressing them, will again be discipline specific

2.5 Incremental

The Incremental Project was a collaboration between the universities of Cambridge and Glasgow.[15] It did not create tools as such, but collected, repurposed, organized, supplemented and repackaged existing guidance on research data management from the two universities. This allowed researchers to find, understand and use the guidance more easily. It also engaged in a programme of researcher training in the form of workshops and seminars on data management issues.

---

12. ICAT-Personal Project Website, URL: http://icatlite.sourceforge.net/
13. Graphviz Website, URL: http://www.graphviz.org/
14. IDMB Project Website, URL: http://www.southamptondata.org/
15. Incremental Project Website, URL: http://www.lib.cam.ac.uk/preservation/incremental/
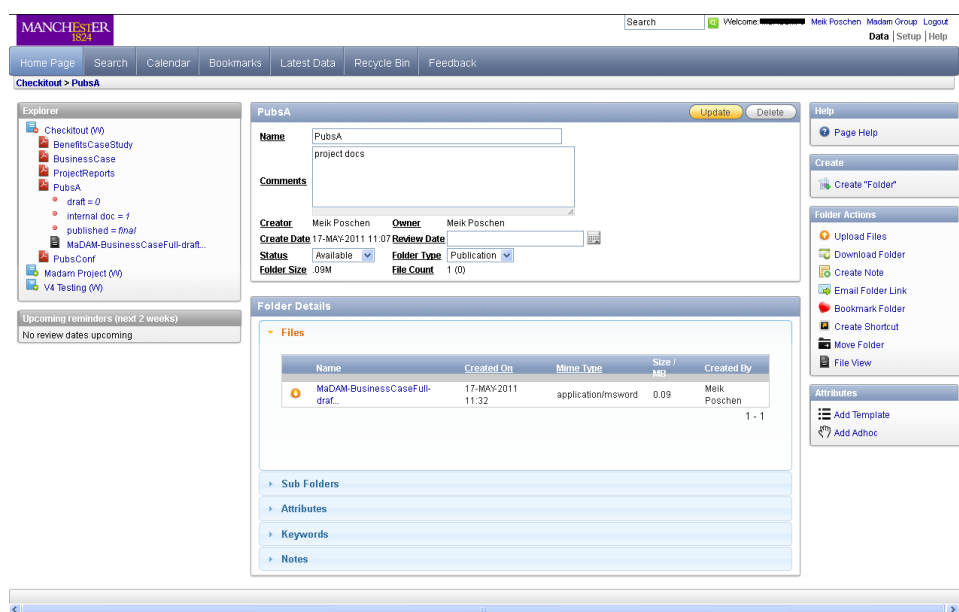
Figure 1: Web-based interface to the MaDAM system

2.6 MaDAM

The purpose of the MaDAM Project was to develop a pilot data management infrastructure for biomedical researchers at the University of Manchester.[16] A major outcome of the project was a Web-based data management environment, also called MaDAM [Pos+11].

The system consists of a Web server running over a relational database management system, and is linked to the university's institutional repository, eScholar, which uses Fedora Commons software. The interface of the MaDAM system (see Figure 1) organizes resources into a hierarchy of directories, with the top level representing projects and lower levels representing experiments, samples, publications and so on. Access controls may be set at both directory and file level, and for both groups and individuals. Files and folders may be tagged with additional information and linked to other files or folders inside or outside the system. The interface includes facilities for searching, browsing and previewing data. It also includes an archiving function, which 'freezes' a dataset, creates a record for it in the institutional repository, and allows the user to decide if the data will remain private, become public after an embargo period, or be made immediately available.

MaDAM also includes its own integrated data management planning tool (eDMP), which is linked to the University of Manchester's Research Information Management System for autocompletion of various entries.

The MaDAM system is being further developed and rolled out at an institutional level by the MiSS Project.[17]

---

16. MaDAM Project Website, URL: http://www.library.manchester.ac.uk/aboutus/projects/madam/
17. MiSS Project Website, URL: http://www.miss.manchester.ac.uk/

Figure 2: DaaS database browse and simple search interface

2.7 PEG-BOARD

The PEG-BOARD Project sought to improve the management of palæoclimate data and simulated climate data held at the University of Bristol.[18] It created data management plans and policies, a scheme of appropriate metadata for managing the data, and improved the backend infrastructure for the BRIDGE research group's software suite.

2.8 SUDAMIH

The SUDAMIH Project produced tools and provided training to support data management in the humanities at the University of Oxford.[19] The main software deliverable was a pilot 'Database as a Service' product consisting of five modules [Wil11].

1. A core administrative interface for managing PostgreSQL relational databases.

2. A conversion utility for importing Access databases and comma-separated value (CSV) files into the system.

3. A graphical tool for creating or modifying database structures and relationships. The structures thus created are portable to other relational database management systems.

4. A graphical tool for building form-based interfaces to the data.

5. A tool to assist users with creating advanced SQL queries.

The first of these is shown in Figure 2. The software is being developed into a cloud-based shared service by the VIDaaS Project, with funding from the Universities Modernization Fund.[20]

---

18. PEG-BOARD Project, URL: http://www.paleo.bris.ac.uk/projects/peg-board/
19. SUDAMIH Project Website, URL: http://sudamih.oucs.ox.ac.uk/
20. VIDaaS Project Website, URL: http://vidaas.oucs.ox.ac.uk/

3 Research data management planning projects

3.1 DMBI

The primary purpose of the DMBI project was to overhaul the data management system at the John Innes Centre so that it better suited the workflows of biomedical researchers while also supporting digital curation activity.[21] The new system assembled by the project uses the OMERO bio-image repository software and OME-TIFF file format developed by Open Microscopy Environment (OME),[22] alongside a MediaWiki installation for use as a knowledge management environment and electronic laboratory notebook.

3.2 DMP-ESRC

The DMP-ESRC project was lead by the UK Data Archive for the purpose of providing best practice guidance, support and training.[23] Highlights from among the project outputs include a new edition of *Managing and Sharing Data* [Van+11] (with chapters on data management planning, documentation, formatting, storage, and legal and ethical issues), a set of recommendations for research centres and programmes [Hor+11] and an activity-based tool for costing data management [DMP11].

3.3 DMP-MRC

The DMP-MRC Project was lead by STFC, with the aim of propagating good data management practice in the field of medical research.[24] It examined current practice and the expectations of the Medical Research Council, and used these to formulate best practice, construct template data management plans, and recommend a roadmap for the future.

3.4 HALOGEN

The HALOGEN Project at the University of Leicester was set up to develop a research database supporting researchers working on the cross-disciplinary Roots of the British project.[25] The database was implemented in MySQL and integrated four major datasets, each in a different format and from a different disciplinary perspective.

3.5 MRD-GW

The MRD-GW Project at the University of Glasgow did not produce tools, but a report on how 'Big Science' data are currently managed and recommendations for future action [GCW11].[26]

---

21. DMBI Project Web page, URL: http://dmbi.nbi.bbsrc.ac.uk/
22. OME Website, URL: http://www.openmicroscopy.org/
23. DMP-ESRC Project Web page, URL: http://www.data-archive.ac.uk/create-manage/projects/jisc-dmp
24. DMP-MRC Project page on the JISC Website, URL: http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmp/mrc.aspx
25. Halogen Project Website, URL: http://www2.le.ac.uk/offices/itservices/resources/cs/pso/project-websites/halogen
26. MRD-GW Project Web page, URL: http://purl.org/nxg/projects/mrd-gw

4 CITING, LINKING, INTEGRATING AND PUBLISHING RESEARCH DATA

4.1 ACRID

The ACRID Project was a collaboration between the University of East Anglia, STFC and the Met Office that developed an approach to publishing climate data held by the former institution.[27] The project

- chose to normalize the datasets to Climate Science Modelling Language (CSML);[28]

- set up a two-stage infrastructure for ingesting data from climate monitoring stations (i.e. with a staging area for raw data and a primary area for normalized data and subsequent refinements, derivations and metadata); and

- used OAI-ORE Resource Maps to provide machine-readable landing pages for datasets, that double as indices of the (RDF-formatted) data and metadata contained within [OAI08b].

4.2 Data Linking with Knowledge-Blogging

The Data Linking with Knowledge-Blogging Project was an initiative of the universities of Newcastle and Manchester to adapt the WordPress blogging platform for the purposes of scholarly publishing.[29] The workflow envisioned for these 'Knowledge Blogs' is as follows. Authors create a paper in the form of a blog post, either directly on the Knowledge Blog, or indirectly, with the title and (optionally) the abstract in a Knowledge Blog post that links to the full paper on an author's blog. Reviewers are selected, and they write their reviews also as blog posts, either on the Knowledge Blog or their own blog, with a link at the start to the paper they are reviewing. The Knowledge Blog uses pingback or trackback technology to ensure the reviews appear as comments to the paper [Wor, s.v. 'Managing Comments']. When all the reviews are complete, the authors edit the paper to address the concerns of the reviewers and then contact the editor. The editor checks that the paper is suitable in the light of the reviews and subsequent edits, and may make additional change requests through the blog comments section. Once satisfied, the editor tags the paper as 'reviewed'.

The project produced several WordPress plugins to assist in the publication of scholarly material in blog posts.

- The MathJax-Latex plugin enables authors to enter mathematical equations in LaTeX format, and have them displayed as vector graphics (i.e. using Web fonts rather than as bitmaps).

- The Knowledge Blog Table of Contents plugin allows one to create a bullet list of papers in a given category, along with their authors, as an analogue to a journal issue's table of contents.

- The KCite plugin transforms marked-up CrossRef DOIs and PubMed identifiers into formatted bibliographic citations.

---

27. ACRID Project Website, URL: http://www.cru.uea.ac.uk/cru/projects/acrid/
28. CSML Website, URL: http://csml.badc.rl.ac.uk/
29. Data Linking with Knowledge-Blogging Project Website, URL: http://knowledgeblog.org/

- The Knowledge Blog Post Metadata plugin allows one to expose bibliographic metadata about a post as `<meta>` tags in the HTML head section.

Some other existing plugins were also used, including Co-Authors Plus (enable multiple authorship), COinS Metadata Exposer, Edit Flow (for the editorial workflow), ePub Export, Post Revision Display (for exposing all post-publication versions), SyntaxHighlighter Evolved and WP Post to PDF [Lor+11].

4.3 Dryad UK

The Dryad UK Project was a collaboration between the University of Oxford, the British Library and others to set up a UK mirror of the evolutionary biology and ecology data repository, Dryad.[30,31] As well as setting up the mirror and adding seven UK publishers to the list of Dryad partners (and who therefore might use Dryad as a data archive for some of their journals), the project also mapped Dryad metadata to RDF and developed tools to support the MIIDI metadata standard (Minimum Information about an Infectious Disease Investigation).[32]

4.4 FISH.Link

The FISH.Link Project was a partner to the FISHnet Project (see Section 2.2) and involved the University of Manchester, the Freshwater Biological Association, the Centre for e-Research at King's College London, and the Queen Mary University of London.[33] The project produced a plugin for XLWrap,[34] provisionally called WaterColumn, to support the publication of data as RDF, thus supporting the process of moving data into the Green category in the Freshwater*Life* repository. WaterColumn can take a spreadsheet from the repository, convert it (if necessary) to Excel and insert metadata selection boxes for each column, populated with options drawn from a controlled vocabulary, then deposit the spreadsheet back to the repository. The intention is for a domain expert to retrieve the spreadsheet, select the appropriate options for each column and, after the selections have been checked by a reviewer, return the spreadsheet to the repository. WaterColumn can then generate an appropriate mapping file that XLWrap can use to convert the data to RDF.

4.5 SageCite

The SageCite Project was a collaboration between UKOLN (University of Bath), the University of Manchester and the British Library to develop a citation framework appropriate to the Sage Commons, ensuring that not only can publications be cited, but also the underlying data and workflows.[35] Among other things, the project produced a plugin for workflow management tool Taverna,[36] allowing one to automate the registration of a DataCite DOI for the output data of a workflow.

---

30. Dryad UK Project wiki, URL: http://wiki.datadryad.org/DryadUK
31. Dryad Website, URL: http://datadryad.org/
32. MIIDI Website, URL: http://www.miidi.org/
33. FISH.Link Project Website, URL: http://www.fishlinkonline.org/
34. XLWrap Web page, URL: http://xlwrap.sourceforge.net/
35. SageCite Project blog, URL: http://blogs.ukoln.ac.uk/sagecite/
36. Taverna Website, URL: http://www.taverna.org.uk/

4.6 SPQR

The SPQR Project, involving King's College London, the University of Edinburgh and Humboldt University Berlin, investigated the possibility of converting data from Classics research into RDF to facilitate their integration.[37] The project did not create any new tools, but did assess several existing tools for their suitability [SPQ11b]. Among those eventually selected were D2R Server (for converting relational databases to RDF),[38] Apache Jena (as a Semantic Web programming framework),[39] AllegroGraph (a triple store),[40] and the Pellet OWL2 reasoner (for inferences) [SPQ11a].[41]

4.7 Webtracks

The Webtracks Project, a collaboration between the University of Southampton and STFC, built on the work of the earlier CLADDIER and StoreLink projects to further develop a protocol for inter-repository communication known as the InteRCom protocol [Mat+10].[42] The aim is for the protocol to support secure communications between data repositories, publication repositories and open science notebooks in order to maintain the linkages between the working records of research activities, released data sets and scholarly publications. The protocol does not appear to have been released at the time of writing.

4.8 XYZ

The XYZ Project was lead by the University of Cambridge, with the aim of developing a new workflow for publishing data alongside journal papers.[43] One of the approaches explored by the project was embedding data directly within the Web pages of journal papers, using RDF and microformats in a profile of HTML known as Scholarly HTML [Sef11].

5 Research Data Management Training Materials

The training materials produced by the five projects in this strand – CAiRO,[44] DataTrain,[45] DATUM for Health,[46] DMTpsych,[47] and Research Data MANTRA[48] – were tailored to specific disciplines and institutions, but nevertheless may be of use to students and

---

37. SPQR Project Website, URL: http://spqr.cerch.kcl.ac.uk/

38. D2R Server Web page, URL: http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/

39. Apache Jena Webpage, URL: http://incubator.apache.org/jena/

40. AllegroGraph Web page, URL: http://www.franz.com/agraph/allegrograph/

41. Pellet Web page, URL: http://clarkparsia.com/pellet/

42. Webtracks Project blog, URL: http://webtracks.jiscinvolve.org/; StoreLink Project Web page, URL: https://sites.google.com/a/staffmail.ed.ac.uk/storelink/; CLADDIER Project wiki (archived), URL: http://web.archive.org/web/20100503183719/http://claddier.badc.ac.uk/trac

43. XYZ Project blog, URL: http://projectxyz.wordpress.com/

44. CAiRO Project Website, URL: http://www.projectcairo.org/

45. DataTrain Project Website, URL: http://www.lib.cam.ac.uk/preservation/datatrain/

46. DATUM for Health Project Website, URL: http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/

47. DMTpsych Project Website, URL: http://www.dmtpsych.york.ac.uk/

48. Research Data MANTRA Project Website, URL: http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/mantra

researchers in the University of Bath's Department of Mechanical Engineering in understanding the need for and process of data management planning.

The Research Data MANTRA course is aimed at both PhD students and those planning research projects.[49] While it was written for students and researchers in the fields of geosciences, social and political sciences and clinical psychology, it is for the most part discipline-agnostic. It conveys both the issues and practices to tackle them in an accessible, multimedia fashion.

Project CAiRO's Managing Creative Arts Research Data (MCARD) training module contains concise textual introductions to several important topics.[50]

- In Unit 2 (Creating), the section on planning to make research data is generically applicable. The section on model release could be relevant in cases where, for example, engineers are filmed to record their design behaviour.

- Unit 3 (Managing) gives good background information about digital curation, though there are some areas where REDm-MED would wish to give more specific advice (e.g. regarding file naming conventions).

The DataTrain Project prepared training modules for both social anthropology PhD students and archæology postgraduates. Within the social anthropology course,[51] the sections of most generic relevance (particularly to resedarchers engaged in qualitative research) are the sections of data creation, capture and organization and back-ups from the Basic Module, the Writing Up Module and the list of useful software. Within the archæology course,[52] the most useful modules are Module 2 (Data Lifecycles and Management Plans), Module 3 (Working with Digital Data), Module 4 (Rights and Digital Data), and, for PhD students, Module 5 (E-Theses and Supplementary Digital Data).

The third session of the DATUM for Health training course (Problems and Practical Strategies and Solutions) may be of interest for its guidance on handling confidentiality, sensitive data and security.[53]

6 Conclusions

The Managing Research Data programme has produced a handful of tools that could be used by researchers or projects on an individual basis. ICAT-Personal could be used at the research group level in those areas characterized by the application of a standard data processing workflow on different sets of raw data. For researchers dealing with relational databases, the VIDaaS tool may make a friendly alternative to using desktop database applications such as Microsoft Access. The DMP-ESRC activity-based costing tool could be of use to those factoring data management costs into bids for funding.

---

49. Research Data MANTRA course, URL: http://datalib.edina.ac.uk/mantra/
50. MCARD Training Module, URL: http://www.projectcairo.org/node/9
51. DataTrain Social Anthropology training modules, URL: http://www.lib.cam.ac.uk/dataman/datatrain/socanthintro.html
52. DataTrain Archæology training modules, URL: http://archaeologydataservice.ac.uk/learning/DataTrain
53. DATUM for Health training course, URL: http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/health/materials/

redm2rep120110ab10.pdf

From a whole-department perspective, the tools of greatest interest are probably the MaDAM and DataFlow environments; currently it seems the former is more polished, but the latter is more flexible in terms of the number of interfaces. A private Knowledge Blog could be considered as an alternative platform for drafting research reports or journal articles, but the system does not appear to have much to specifically support research data as distinct from other types of 'attachment'. The DMBI Project provides some validation of ERIM's approach of using a wiki as a collaborative environment for recording research.

References

[BPW11]   M Brown, O Parchment & W White (2011-08-31). *Institutional Data Management Blueprint*. IDMB Project Document. University of Southampton. URL: http://eprints.soton.ac.uk/196241/.

[CC]   *About CC0 — 'No Rights Reserved'*. Creative Commons. URL: http://creativecommons.org/about/cc0 (2012-02-14).

[DB12]   M Darlington & A Ball (2012-02-10). *Research Data Management Plan Requirements Specification for the Department of Mechanical Engineering, University of Bath*. REDm-MED Project Document redm1rep120105mjd10. URL: http://opus.bath.ac.uk/28040/.

[DCM10]   DCMI Usage Board (2010-10-11). *DCMI Metadata Terms*. DCMI Recommendation. Dublin Core Metadata Initiative. URL: http://dublincore.org/documents/2010/10/11/dcmi-terms/.

[DMP11]   DMP-ESRC Project (2011). *Costing Research Data Management in the Social Sciences*. UK Data Archive. URL: http://www.data-archive.ac.uk/media/257647/ukda_jiscdmcosting.pdf (2012-02-17).

[GCW11]   N Gray, T Carozzi & G Woan (2011-06-29). *Managing Research Data – Gravitational Waves: Final Report*. LIGO Document LIGO-P1000188-v10. URL: http://purl.org/nxg/projects/mrd-gw/report.

[Haf10a]   M Haft (2010-11-15). 'A Simple Traffic Light System for Data'. *FISHnet Blog*. URL: http://www.fishnetonline.org/home/-/blogs/a-simple-traffic-light-system-for-data.

[Haf10b]   M Haft (2010-07-23). 'FISHnet, JIRA and Being Agile'. *FISHnet Blog*. URL: http://www.fishnetonline.org/home/-/blogs/fishnet-jira-and-being-agile.

[Haf10c]   M Haft (2010-05-11). 'The FISHNet Server-Side Technology Stack'. *FISHnet Blog*. URL: http://www.fishnetonline.org/home/-/blogs/the-fishnet-server-side-technology-stack.

[Hor+11]   L Horton et al. (2011-03-28). *Data Management Recommendations for Research Centres and Programmes*. UK Data Archive: Colchester, UK. URL: http://www.data-archive.ac.uk/media/257765/ukda_datamanagementrecommendations_centresprogrammes.pdf.

[I2S11]   I2S2 Project (2011-04-07). *I2S2 Idealised Sceintific Research Activity Lifecycle Model*. URL: http://www.ukoln.ac.uk/projects/I2S2/documents/I2S2-ResearchActivityLifecycleModel-110407.pdf.

[Jon]   R Jones. *SWORD V2 Specifications*. URL: http://swordapp.org/sword-v2/sword-v2-specifications/ (2012-02-14).

[Lor+11]   P Lord et al. (2011-06-07). 'The Ontogenesis Knowledgeblog: Lightweight Semantic Publishing'. *Knowledge Blog*. URL: http://knowledgeblog.org/128.

[Mat+10]   B Matthews et al. (2010). 'A Protocol for Exchanging Scientific Citations'. In: *Proceedings of the 5th IEEE International Conference on e-Science.* (Oxford, UK, 9th–11th Dec. 2009). IEEE Computer Society: Washington, DC, 171–177. ISBN: 978-0-7695-3877-8. DOI: 10.1109/e-Science.2009.32.

[OAI08a]   C Lagoze et al. (eds.) (2008-12-07). *Open Archives Initiative Protocol for Metadata Harvesting.* URL: http://www.openarchives.org/OAI/openarchivesprotocol.html.

[OAI08b]   *ORE Specifications and User Guides* (2008-10-17). *Table of Contents* (2008-10-17). Open Archives Initiative. URL: http://www.openarchives.org/ore/1.0/toc.html.

[Pos+11]   M Poschen et al. (2011). 'Development of a Pilot Data Management Infrastructure for Biomedical Researchers at University of Manchester: Approach, Findings, Challenges and Outlook of the MaDAM Project'. In: *7th International Digital Curation Conference.* (Bristol, UK, 5th–7th Dec. 2011). URL: http://www.merc.ac.uk/sites/default/files/MaDAM-IDCC2011-PracticeTrack-Paper-final.pdf.

[Sef11]   P Sefton (ed.) (2011-05-03). *Scholarly HTML Core.* URL: http://scholarlyhtml.org/2011/05/03/scholarly-html-core-3/.

[SPQ11a]   *Assessing linked data tools for SPQR* (2011). SPQR Project. URL: http://spqr.cerch.kcl.ac.uk/?page_id=112 (2012-02-20).

[SPQ11b]   *Linked Data Tools* (2011). SPQR Project. URL: http://spqr.cerch.kcl.ac.uk/?page_id=94 (2012-02-20).

[Tak10]   K Takeda (2010-09-15). *Initial Findings Report.* IDMB Project Document. University of Southampton. URL: http://eprints.soton.ac.uk/195155/.

[Van+11]   V Van den Eynden et al. (2011-05). *Managing and Sharing Data: Best Practice for Researchers.* 3rd ed. UK Data Archive: Colchester, UK. ISBN: 1-904059-78-3. URL: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf.

[W3C04]   F Manola & E Miller (eds.) (2004-02-10). *RDF Primer.* W3C Recommendation. World Wide Web Consortium. URL: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

[W3C08]   E Prud'hommeaux & A Seaborne (eds.) (2008-01-15). *SPARQL Query Language for RDF.* W3C Recommendation. World Wide Web Consortium. URL: http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/.

[Wil11]   JAJ Wilson (2011-05-10). *SUDAMIH Final Report.* SUDAMIH Project Document. University of Oxford. URL: http://sudamih.oucs.ox.ac.uk/docs/Sudamih_FinalReport_v1.0.pdf.

[Wil12]   P Willett (2012-02-10). *BagIt File Packaging Format.* URL: https://confluence.ucop.edu/display/Curation/BagIt.

[Wor]   WordPress. 'Introduction to Blogging'. In: *WordPress Codex.* URL: http://codex.wordpress.org/Introduction_to_Blogging (2012-02-17).

[YM11]   E Yang & B Matthews (2011-03-31). *Pilot Implementation.* I2S2 Project Deliverable 3.3. University of Bath & University of Southampton. URL: http://www.ukoln.ac.uk/projects/I2S2/documents/I2S2-WP3-D3.3-PilotImplementation.pdf.