



Research Data Management in Higher Education

Data For Life: Digital Preservation for Health Sciences, Northumbria University, 26 May 2011

Simon Hodson, JISC Programme Manager, s.hodson@jisc.ac.uk



- Drivers: funders policies/requirements; UKRIO, Fol, journals.
- Responses: Edinburgh's institutional policy; JISCMRD Programme
- Benefits and implications.
- Outline of new JISCMRD Programme.

- Research funders' policies form an important part of the research data ecology.
- In common with international developments requirements are becoming increasingly exacting.
- Many policy statements reference the OECD Principles and Guidelines for Access to Research Data from Public Funding: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- **NSF** recently added the requirement of a data management plan to grant proposals: <http://www.arl.org/rtl/eresearch/escien/nsf/index.shtml>
- Health Research Funders' 'Joint Statement of Purpose: **Sharing research data to improve public health**': <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm>
 - making research data sets available to investigators beyond the original research team in a timely and responsible manner, subject to appropriate safeguards, will generate three key benefits:
 - faster progress in improving health
 - better value for money
 - higher quality science

- Joint **RCUK** 'Common Principles' now released:
<http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>
- **EPSRC** shortly to release Policy Framework stating expectations concerning the Management of and Access to EPSRC-funded Research Data. ***Appears to place responsibility primarily with departments and centres to show they can manage and preserve data to adequate standards.***
- New **MRC** policies on research data **management** and **sharing** being prepared; tested and refined; guidance produced as part of a JISC funded project.
- **BBSRC** Statement, April 2007, updated June 2010:
<http://www.bbsrc.ac.uk/web/FILES/Policies/data-sharing-policy.pdf>
 - Requires statement on data sharing.
- New **ESRC** policy now in vigour: <http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>
 - Introduces the requirement of a data management and sharing statement (J-eS) and a data management and sharing plan as part of the grant submission

- UKRIO Code of Practice for Research: Promoting good practice and preventing misconduct, September 2009:
<http://www.ukrio.org/resources/UKRIO%20Code%20of%20Practice%20for%20Research.pdf>
- Data management planning is an essential part of research design [3.4.1.c; also 3.12.6]
- **Section 3.12 covers collection AND RETENTION** of research data.
- Organisations and researchers should ensure that research data relating to publications is available for discussion with other researchers, subject to any existing agreements on confidentiality. [3.12.1]
- Organisations should have in place procedures, resources (including physical space) and administrative support to assist researchers in the accurate and efficient collection of data and its storage in a secure and accessible form. [3.12.5]
- Due regard to privacy, confidentiality and ethical issues.
- But presumption to share: **research integrity requires addressing these issues in order to make data as sharable as possible.**

- Research data can be subject to Freedom of Information / Environmental Information requests: UEA and Queen's University Belfast cases.
- JISC Q&A on 'Freedom of Information and Research Data':
<http://www.jisc.ac.uk/publications/programmerelated/2010/foiresearchdata.aspx>
- Indicative research by Chris Rusbridge on number of FoI requests for research data: sample of 21 Universities, received total of 40 FoI requests for research data from 2007-10.
 - Wide variance in distribution 12 universities received 0; 1 received 8; another 9.
 - All but six were from 2009 and 2010.

- **INSPIRE:** framework for sharing metadata and providing access to geospatial data.
 - Universities **are** considered public authorities and are subject to INSPIRE.
 - Metadata to be made available according to a standard and a schedule.
 - Annex 1 and 2 cover fundamental geospatial reference data (topology, boundaries etc): metadata to be available by 24 Dec 2010.
 - Annex 3 covers 'human geography', including 'Human health and safety': 'Geographical distribution of dominance of pathologies (allergies, cancers, respiratory diseases, etc.), information indicating the effect on health (biomarkers, decline of fertility, epidemics) or well-being of humans (fatigue, stress, etc.) linked directly (air pollution, chemicals, depletion of the ozone layer, noise, etc.) or indirectly (food, genetically modified organisms, etc.) to the quality of the environment.'
 - Annex 3 metadata to be available by 24 Dec 2013.
 - Briefing on INSPIRE being produced as part of JISC Geospatial Programme of activity.

■ **Nature:** http://www.nature.com/authors/editorial_policies/availability.html

- Requires as a condition of publication that authors ‘make materials, data and associated protocols promptly available to readers without undue qualifications’, ideally via a public repository.
- Lists community endorsed repositories for which deposit is mandatory and recommends other public repositories at ‘earlier stage of development’ short of having received full community endorsement.

■ **‘Dryad’ Consortium of Evolutionary Biology Journals:** <http://datadryad.org>

- Joint declarations in *American Naturalist*, *Evolution*, the *Journal of Evolutionary Biology*, *Molecular Ecology*, *Heredity*, and other key journals in evolution and ecology...
- This journal requires, as a condition for publication, that data supporting the results in the paper should be archived in an appropriate public archive...
- Named existing data repositories for specific data types (e.g. Genbank) and the Dryad repository for all other data.
- Dryad now has 70 associated journals.
- JISC-funded Dryad-UK Project looking at further expansion (**infectious disease**) and business model.

Response: University of Edinburgh's RDM Policy

- Aspirational 10-point policy: <http://www.ed.ac.uk/is/research-data-policy>
- 1. Research data will be managed to the highest standards throughout the research data lifecycle **as part of the University's commitment to research excellence.**
- 2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with **Principal Investigators** (PIs).
- 3. All new research proposals [from date of adoption] must include **research data management plans** or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
- 4. **The University will provide training, support, advice** and where appropriate guidelines and templates for the research data management and research data management plans.
- 5. **The University will provide mechanisms and services** for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.

Response: University of Edinburgh's RDM Policy (2)

6. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under **appropriate safeguards**.
8. **The legitimate interests of the subjects of research data must be protected.**
9. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.
10. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without **retaining the rights to make the data openly available for re-use**, unless this is a condition of funding.

Drivers

- Good data management and curation is **essential** for research integrity.
- **Research funders** increasingly insisting that researchers to make data available (the outputs of publicly funded research are a public asset).
- **Scholarly publications** *should* be underpinned/provide access to the underlying data.

Benefits

- There **are identifiable research benefits** from improved data management, including improved efficiency.
- Research benefits of data sharing, large international collaborations and data infrastructures: genomics (e.g. Bermuda and Fort Lauderdale accords); climate science (IPCC data requirements).
- Increasing recognition of the reuse **value** of research data (data integration, metastudies etc).
- **Researchers** attitude to data sharing uneven across disciplines, qualified, conditional but not insuperable.

- First set of projects started 18 months ago, October 2009
- Total of 30 projects.
- Total investment of c.£4.3M from JISC.
- Some are projects are ending imminently; the remainder will finish between now and July.

- **Planning:** There is little institutional support for funders requirements: how do I create an effective data management plan and how do I put it into action?
- **Infrastructure:** There is little effective infrastructure: storage, sharing and management tools are ad hoc.
- **Support and Training:** How well are researchers being prepared for data sharing? Help in meeting funders, institutions requirements?
- **Incentives and Benefits:** Why should I put effort into the data? What recognition and reward will I get for publishing 'my' data? Effective links between publications and data? Effective citation for data and inclusion in metrics?

- Need to improve (and make a case for) data management **planning**.
 - Six projects producing case studies, guidance and model data management plans
- Need to develop RDM **infrastructure** in institutions, build evidence, provide examples to stimulate change.
 - Eight projects gathering requirements, designing infrastructure in the broad sense and implementing; also obtaining cost and benefits information; preparing business cases for project continuity.
 - Two support projects providing help with requirements gathering, obtaining cost-benefits evidence, business cases.
- Need to promote **open publication of data** and develop mechanisms for this.
 - Eight projects exploring means of citing, linking, integrating and publishing data, and demonstrating the benefits.
- Need to develop training materials to improve **capacity and skills** in the sector.
 - Five project designing discipline focussed data management training units for PG courses
 - One support project, guidance and stakeholder engagement.
- Projects completing between April and July 2011
- JISC Website: <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>

- Premise that on the whole, a lack of subject focused training, in data management and sharing, as part of PG and UG studies.
- Projects **to design and pilot (reusable) discipline-focussed training units for postgraduate courses.**
- While such units must necessarily consider generic issues in research data curation, **it is essential that projects collaborate closely with specific, named academic departments within their institution in order to develop effective and useful discipline-focussed training materials.**
- Encourage collaboration between research oriented departments and other bodies in Universities which may have expertise in data and information management; and between Universities and Data Centres.
- Five new projects, completing July 2011.
- DAMSSI Support Project: Led by DCC, with RIN Research Information Handling Working Group (SCONUL, Vitae BAILER etc).

Training (RDMTrain) Projects

DATUM for Health	Northumbria University, DPC, DCC	Health Studies: http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/
CAiRO	University of Bristol, DCC	Performance Studies and related disciplines: http://www.projectcairo.org/
DataTrain	University of Cambridge, ADS	Archaeology and Social Anthropology: http://www.lib.cam.ac.uk/preservation/datatrain/
Research Data MANTRA	University of Edinburgh	Social and Political Sciences, Geographical Sciences, Clinical and Health Psychology: http://bit.ly/RDMantra
RDM in Psychological Sciences	University of York	Psychological Sciences: http://www.dmtpsych.york.ac.uk/
DaMSSI Support Project	DCC / RIN	Support Role, liaison and benchmarking with key stakeholders (Vitae, SCONUL etc).

- To help researchers and institutions respond to funders' policies
- Projects identified in conjunction with funders.
- Plans not just tick box exercise, but to be implemented.
- To understand and to specify in practical terms how effective data management planning can best be enabled within research projects.
 - Case studies of improvement in data management practice.
 - Training and outreach in data management planning.
 - Model data management plans.
- Six projects, completing April – July 2011.

EPSRC	Bath, IdMRC	ERIM Project (E ngineering Research Information Management): http://www.bath.ac.uk/idmrc/erim/
STFC	Glasgow	Data management planning and community solutions in gravitational wave astronomy : http://www.astro.gla.ac.uk/users/norman/projects/mrd-gw/
Wellcome	Leicester	HALOGEN (History, Archaeology, Linguistics, Onomastics, Genetics) - interdisciplinary : http://www2.le.ac.uk/offices/itservices/resources/cs/psd/its-project-portfolio/halogen
ESRC	UKDA	Data management planning for ESRC data-rich investments (social sciences): http://www.data-archive.ac.uk/create-manage/projects/JISC-DMP ; http://ukdaresearchdatamanagement.blogspot.com/
BBSRC	John Innes Centre	DMBI (Data Management for Bio-Imaging – i.e. plant microscopy): http://dmbi.nbi.bbsrc.ac.uk/index.php/Main_Page
MRC	STFC – MRC DSS	MRC Data Management Plans for Patient and Population Cohort Study data sharing : http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmp/mrc.aspx

DMP-ESRC Project

- Run by UK Data Archive: <http://www.data-archive.ac.uk/create-manage/projects/jisc-dmp>
- Study of data management practices in ESRC funded Centres and Programmes.
- Data Management Recommendations for Research Centres and Programmes: http://www.data-archive.ac.uk/media/257765/ukdadatamanagementrecommendations_centresprogrammes.pdf
 - Clear roles and responsibilities; RDM coordinator; Data Inventory; Data Management Resources Library.
 - Recommendations and guidelines on anonymisation, security and backup etc.
- Data Management Costing Tool: http://www.data-archive.ac.uk/media/257647/ukda_jiscdmcosting.pdf

- To provide a pilot projects for the development of **Research Data Management Infrastructure**, in and across HEIs.
- **Approach:**
 - Requirements gathering;
 - Implementing a pilot infrastructure (including support, policies, training);
 - **Cost-benefits analysis and business model for sustainability.**
- To provide the sector as a whole with lessons and examples of **how** a broader data management infrastructure may be constructed.
- Eight projects running from Oct 2009 – March/June 2011.

Infrastructure (RDMI) Projects

Incremental	Cambridge, Glasgow	Focus on training, guidance and support materials; variety of subject areas: http://www.lib.cam.ac.uk/preservation/incremental/
Sudamih	Oxford	Humanities; training and support; database hosting: http://sudamih.oucs.ox.ac.uk/ ; http://blogs.oucs.ox.ac.uk/sudamih/
FISHNet	KCL, Freshwater Biological Association	Freshwater biology: small scale observational data, large scale monitoring data: http://www.fishnetonline.org/home
I2S2	Bath, Southampton, Cambridge, STFC-RAL	Structural science, data management between facilities (Diamond synchrotron, ISIS facility) and lab: http://www.ukoln.ac.uk/projects/I2S2/
Admiral	Oxford	Zoology (includes neutron diffraction data, spreadsheets of lab data, microscopy image data, GPS data): http://imageweb.zoo.ox.ac.uk/wiki/index.php/ADMIRAL
IDMB	Southampton	Institutional Blueprint: Archaeology; Chemistry; Electronics and Computer Science; Engineering Sciences: http://www.southamptondata.org/
MaDAM	Manchester	Life and Biomedical Sciences (electron and standard microscopy; MR imaging neuropsychiatry): http://www.library.manchester.ac.uk/aboutus/projects/madam/
PEG-BOARD	Bristol	Palaeoclimatology modelling data: http://www.paleo.bris.ac.uk/projects/peg-board/

Requirements Analyses

- Data assets of research done in Universities often comprises small, labour intensive data.
- **Storage and curation of data falls, for the most part, falls very far short of good practice.**
- Key JISC reports:
 - Dealing with Data:
http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf
 - Keeping Research Data Safe:
<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>
 - UKRDS Scoping Study:
<http://www.ukrds.ac.uk/resources/>
- JISCMRD Requirements Analyses (current practices, researchers attitudes, requirements).
 - Oxford, **Sudamih**: : <http://sudamih.oucs.ox.ac.uk/>
 - Cambridge, **Incremental**:
<http://www.lib.cam.ac.uk/preservation/incremental/>
 - Southampton, **Institutional Data Management Blueprint**: <http://www.southamptondata.org/>



Sudamih Researcher Requirements Report

Supporting Data Management Infrastructure in the Humanities (Sudamih)

sudamih.oucs.ox.ac.uk

Authors

Dr. James A. J. Wilson
Dr. Muriel Patrick

Affiliation

Oxford University Computing Services

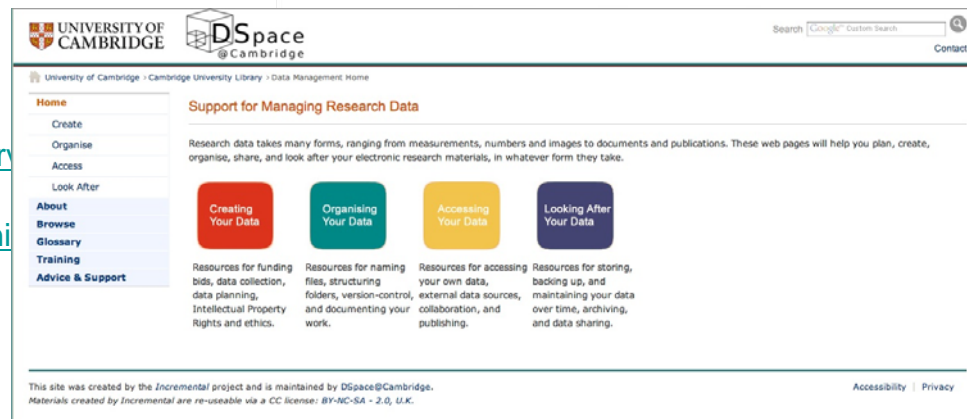
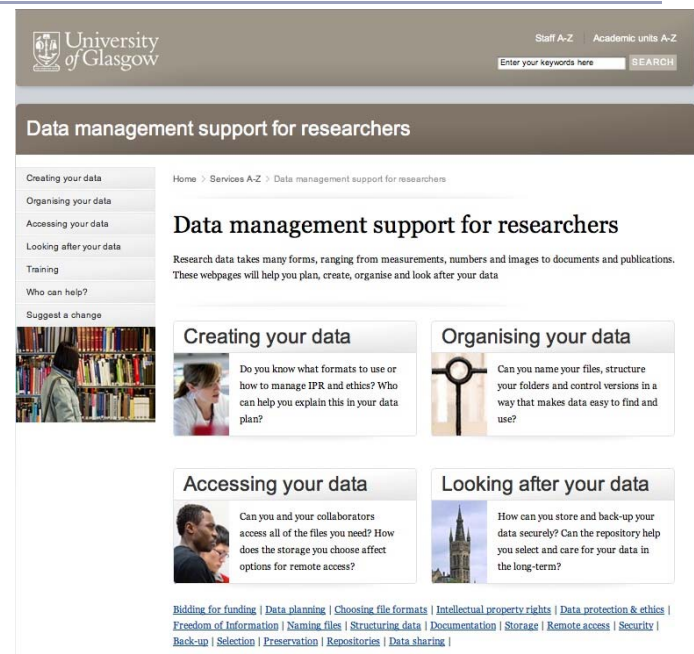
JISC



Incremental

Scoping study and implementation plan
'A pilot project for supporting research data management'

- Incremental Project, collaboration between Glasgow and Cambridge, concentrated on providing guidance and training materials at an institutional level; focus on arts and humanities, social sciences, archaeology, social anthropology:
<http://www.lib.cam.ac.uk/preservation/incremental/index.html>
- Cambridge Website: www.lib.cam.ac.uk/dataman/
- Glasgow Website: www.gla.ac.uk/datamanagement/
- Workshops and Seminars:
<http://www.lib.cam.ac.uk/preservation/incremental/seminars.html>
 - Series at CRASSH covering: ethics, FoI, IPR, new technologies.
 - Series at Glasgow covering: performing arts and archaeology.
- Interviews from Seminars:
 - <http://www.lib.cam.ac.uk/dataman/training.html#Interviews>
 - <http://www.gla.ac.uk/services/datamanagement/training/videos/>
- Incremental Project Blog:
<http://incrementalproject.wordpress.com/>



RDMI Project Outputs: Sudamih


- Sudamih Project:
<http://sudamih.oucs.ox.ac.uk/>
- Oxford Research Data Management Pages (EIDCSR Project):
<http://www.admin.ox.ac.uk/rdm/>
- Reports and Training Materials:
<http://sudamih.oucs.ox.ac.uk/documents.xml>
- Workshops
 - Data Management Training for the Humanities - A Half-Day Workshop
http://sudamih.oucs.ox.ac.uk/training_workshop.xml
 - Research databases in the humanities - where next?
http://sudamih.oucs.ox.ac.uk/databases_workshop.xml
- Prototype **Database as a Service** to be taken up by Oxford Humanities Division and further developed as part of the VIDaaS Project (subject to HEFCE approval).

RESEARCH DATA MANAGEMENT
UAS

[This site](#)
[University of Oxford](#)
[People](#)


[UAS Home](#) > [Research Data Management](#) >

Research Data Management



- » Why manage your data?
- » Data Management Planning
- » Data Backup and Security
- » Data Sharing and Archive
- » Training, Advice & Support


University of Oxford commitment to research data management:



"The University of Oxford is committed to supporting researchers in appropriate curation and preservation of their research data, and where applicable in accordance with the research funders' requirements." NB. Clicking on this link will take you out of the current site) (Source: PRAC ICT Subcommittee)

Good practice in data management is one of the core areas of research integrity, or the responsible conduct of research.

The following diagram provides further insight to some of the stages involved in research data management, and the facilities and services available to help, both within the University and from external providers.



Quick links

- » Data management planning checklist
- » Funder policies
- » Training, advice & support
- » 101 Flyer - 'Managing your research data at The University of Oxford' (916kb)

Find out more

- » UK Data Archive - 'Managing and Sharing Data' (1,188kb)

What's new

- » January 2011 - Wellcome Trust et al: Sharing research data to improve public health: joint statement of purpose (external link)
- » ESRC Research Data Policy - Sep'10 (148kb)
- » 'UK e-Infrastructure report' (pdf - right click to save) - A Review Expert Group, commissioned by BIS, with RCUK taking the lead, has just published its report on 'UK e-Infrastructure'. The e-Science Directors' Forum gave input to the Review. A helpful summary is available at: <http://www.rcuk.ac.uk/escie>

[data management planning checklist](#) | [funder policy](#) | [data management plans](#) | [ethical and legal considerations](#) | [organisation and documentation](#) | [data backup](#) | [data security](#) | [to share or not to share?](#) | [deposit your data](#) | [Training, advice and support](#)

Benefits of better data management

- **Benefits Case Studies and Business Cases:** Synthesis to be released in **June**.
- Benefits achieved upstream of data sharing and publication.
- Benefits and good response to training (Glasgow, Cambridge, Oxford).
- Better practice, security, backup, retrieval through central provision of storage; can be more competitive (Leicester, Manchester, Southampton, Oxford x 2).
- Better practice for research data storage, reduce duplication and redundancy (Manchester, Oxford).
- Greater efficiency, timesaving, improved retrieval of information (Manchester, Oxford, I2S2).
 - I2S2 system reduces data retrieval delay from 24hrs+ to 5-10 minutes.
- Support for researcher workflow, linking between management system and repository for publishing/sharing (Manchester, Oxford, FISHnet, I2S2).
- **Challenges in benchmarking and establishing metrics for benefits: lessons learnt.**

- Support for an implementation of data management planning.
- Systems to support data management through the lifecycle.
- Better understanding of the costs of storage:
 - data appraisal and selection (what data needs to be kept?)
 - data triage for efficient use of storage
- How to recoup costs? Need to understand division of costs between institutions, funders etc.
- Requires PI responsibility and a data manager role.
- Encourage Universities to develop RDM policies like Edinburgh.
- Encourage development of business plans for RDM infrastructures, including storage, systems and support. **A case should be made for the benefits of sustaining and expanding the infrastructure in support of policy.**

What next? New JISCMRD Programme, 2011-13

- Call in late-May / early-June for projects to start by 1 October, up to 18 months; possibly as much as 3M available.
- A: Institutional Research Data Management Infrastructure
 - Projects to **implement** institutional research data management infrastructure, inc. data capture, research data management systems, deposit to repository solutions, institutional policy, business case, provide costing information.
- B: Research Data Management Planning for Projects, Departments or Centres
 - Projects to design **and implement** research data management plans for specific projects/departments; includes supporting systems, tools; provide costing information.
- C: Integrated Research Data Management Planning Tools for Institutions
 - Projects to develop institutional data management planning tools which are integrated with institutional grant costing and research management systems. Such tools must interoperate with, the rules-based DMPonline tool.
- *RDMTraining : data managers; liaison librarians/support roles; researchers.*
- *Infrastructure for Interoperability and Data Publishing (deposit, citation, collaborations between data centres and publishers).*
- **s.hodson@jisc.ac.uk**



Because good research needs good data

- Digital Curation Centre, JISC-funded centre of expertise in digital curation, providing guidance, national coordination: <http://www.dcc.ac.uk>
- DMPonline, planning tool incorporating funders' requirements: <http://dmponline.dcc.ac.uk/>
- **CARDIO** (Collaborative Assessment of Research Data Infrastructure and Objectives), institutional/departmental self assessment tool, designed for research data managers: <http://cardio.dcc.ac.uk>
- Data Asset Framework: <http://www.data-audit.eu/>
- How-To Guides, 'Appraisal' and 'Licensing' so far: <http://www.dcc.ac.uk/resources/how-guides>
- Outreach: Roadshows, Research Data Management Forum: <http://www.dcc.ac.uk/events>
- International Digital Curation Conference (Bristol, Bristol, 5-7 Dec 2011)
- Training: <http://www.dcc.ac.uk/training>

DMPonline

The  | D | C | C Data Management Planning Tool



Funded by:

JISC

- Research Data Management Forum (Warwick, theme TBA, 2-3 November): <http://www.dcc.ac.uk/events/research-data-management-forum>
- DCC Roadshows (next Glasgow, 22-24 June; then Newcastle, Oxford, Brighton, Cardiff and Cambridge): <http://www.dcc.ac.uk/events/data-management-roadshows>
- 7th International Digital Curation Conference, Bristol, 5-7 Dec 2011: <http://www.dcc.ac.uk/events/idcc11>

- 6-7M University Modernisation Fund (UMF) investment in research data infrastructure:
 - c.1.5M to JANET and Eduserv for cloud and brokerage.
 - c.1.5M to expand DCC activity.
 - c.3-4M for RDM tools to work with the cloud offer: closed Call to JISCMRD Projects led to four projects being recommended for funding, subject to HEFCE approval.

- JISC Website: <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>
- Programme Netvibes: <http://www.netvibes.com/jiscmrd#General>
- Programme Blog: <http://researchdata.jiscinvolve.org/>
- Community discussion list: RESEARCH-DATAMAN@JISCMAIL.AC.UK
- List for programme management information, message to/between projects: JISCMRD@JISCMAIL.AC.UK
- Twitter Tag: #jiscmrd
- **s.hodson@jisc.ac.uk**

THANK YOU
QUESTIONS?