



# Science and Technology Resources on the Internet

## Selected Internet Resources on Digital Research Data Curation

### **Brian Westra**

Lorry I. Lokey Science Data Services Librarian  
University of Oregon  
Eugene, Oregon  
[bwestra@uoregon.edu](mailto:bwestra@uoregon.edu)

### **Marisa Ramirez**

Digital Repository Librarian  
California Polytechnic State University  
San Luis Obispo, California  
[mramir14@calpoly.edu](mailto:mramir14@calpoly.edu)

### **Susan Wells Parham**

Research Data Project Librarian  
Georgia Tech Library & Information Center  
Atlanta, Georgia  
[susan.parham@gatech.edu](mailto:susan.parham@gatech.edu)

### **Jeanine Marie Scaramozzino**

College of Science and Mathematics Librarian  
School of Education Librarian  
California Polytechnic State University  
San Luis Obispo, California  
[jscaramo@calpoly.edu](mailto:jscaramo@calpoly.edu)

---

Copyright 2010, Brian Westra, Marisa Ramirez, Susan Wells Parham, and Jeanine Marie Scaramozzino. Used with permission.

[Introduction](#)

[Scope and Methods](#)

[Policies, Best Practices, and Guidelines](#)

[Roles, Services, and Skills](#)

[Sustainability and Cost Projections](#)

[Directories of Data Repositories](#)

[Metadata Standards](#)

[Software and Middleware](#)

[Assessment Tools](#)

[Associations and Organizations](#)

[Open Access Journals](#)

[E-mail Lists](#)

Networked, data-intensive computational approaches to science play an increasingly important role across research disciplines, and this technology-rich environment alters both the content and modes of scholarly communication ([Wright et al. 2007](#)). The sheer volume of digital data produced in the sciences is staggering, presenting challenges to researchers and to publishers, funding agencies, and others both within and external to the academic community ([Borgman, Wallis, and Enyedy 2007](#)).

The curation of digital research data and the development of related infrastructure are of great significance to the research community, as evinced by the National Science Foundation's DataNet Program. The program will ultimately fund five large-scale projects which will include new types of research organizations (DataNet Partners) to focus their research efforts in these areas. The curation of research data is defined as "stewardship that adds value through the provision of context and linkage: placing emphasis on publishing data in ways that ease re-use and promoting accountability and integration" ([Rusbridge et al. 2005](#)).

According to Sayeed Choudhury, the implications of research data curation for libraries include: "Data as collections; data as services; librarians as data scientists; data centers as the new library stacks" ([Choudhury 2009](#)). As such, research libraries are developing services to support the intake, preservation and reuse of this digital content, and exploring new roles for libraries and librarians.

A growing number of institutions have undertaken "collaborative action by cross-section partnerships of academe, government, industry and others" to develop new data curation resources, including case studies, standards and tools, policies, and training ([Gold 2010](#)). The resources listed below are reflective of the current state of knowledge, which will likely undergo significant changes as curation services and technology continue to evolve.

## Scope and Methods

In order to present a webliography of reasonable scope and length, the authors focused on resources applicable to the broader topic of digital research data curation as they relate to the natural sciences. Materials primarily or solely devoted to medical informatics, social sciences, and the humanities were not included. However, it should be noted that a number of the resources presented here are also applicable to research data curation in disciplines other than the sciences -- for example, data repository software may be as useful to the social scientist as it is to a researcher in ecology. Additional scope specificity, when necessary, is provided in respective section listings below.

The identification of resources occurred during the course of the authors' work and research in data curation; through participation in conferences and workshops; via professional contacts; during participation in and searches of e-mail lists, blogs/RSS feeds, wikis, and other social media; and from searches of open access publications of associations, agencies, and commercial publishers.

## Policies, Best Practices, and Guidelines

This section of materials published from 2005 to the present includes documents from associations and organizations listed in the Associations section. (i.e. working documents, wikis, and online presentations).

### **British Library Digital Preservation Strategy**

<http://www.british-library.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresstrat.pdf>

This document sets out the strategy the British Library intends to use to achieve its goal of preserving all of its digital collections in a secure digital repository by 2016. Released 2006.

### **Digital Curation Centre (DCC) Curation Reference Manual**

<http://www.dcc.ac.uk/resources/curation-reference-manual>

This manual includes completed chapters and chapters in production, and is a peer-reviewed, community-driven project to provide advice, in-depth information, and criticism on current digital curation techniques and best practices. No release date, regularly updated.

### **Digital Preservation Policies Study**

<http://www.jisc.ac.uk/publications/reports/2008/jiscpolicyfinalreport.aspx>

This Joint Information Systems Committee (JISC)-funded study provides an outline model for digital preservation policies and the role that digital preservation can play in supporting and

delivering key strategies for higher education institutions. JISC created two tools in the study: (1) a model/framework for digital preservation policy and implementation clauses based on examination of existing digital preservation policies; and (2) a series of mappings of digital preservation links to other key institutional strategies in UK universities and colleges. Released October 2008.

### **Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century**

<http://www.nsf.gov/pubs/2005/nsb0540/>

This National Science Foundation-funded report from the National Science Board includes a discussion of the issues surrounding digital data collection policies, and roles and responsibilities of both individuals and institutions. Appendix C includes a list of the policies on data sharing and archiving that were current at the time of publication. Released 2005.

### **MIT Libraries -- Data Management and Publishing Guide**

<http://libraries.mit.edu/guides/subjects/data-management/index.html>

This guide, primarily written for researchers, includes an introduction to research data management, a data planning checklist, guidance on creating a data management plan, and links to additional guides on research data management and curation. No release date, regularly updated.

### **National Library of Australia Preservation Policy**

<http://www.nla.gov.au/policy/pres.html>

The policy defines the National Library's preservation responsibilities, and provides guidance to library staff engaged in making decisions and undertaking other activities that may impact collections. Specific activities reviewed include: acquisition, safe storage, processing and maintenance, conservation, preservation, and electronic collections. Reviewed August 2009.

### **OCLC Digital Archive Preservation Policy**

<http://www.oclc.org/Support/documentation/digitalarchive/preservationpolicy.pdf>

This document discusses preservation strategies and action plans, including details on data integrity and digital archive risk assessment. Released 2006.

### **Research Data Management and Publishing Support at Cornell**

<https://confluence.cornell.edu/display/datasupp/Home>

This wiki site serves as a directory to research data services at Cornell and includes details on data management planning, storage and backup services, metadata, intellectual property, data publication, and more. No release date, regularly updated.

### **Research Information Network's Stewardship of Digital Research Data**

<http://www.rin.ac.uk/our-work/data-management-and-curation/stewardship-digital-research-data-principles-and-guidelines>

This document outlines a framework of principles and guidelines for research institutions and funders, research data curators, professional societies, and publishers. Released 2008.

### **University of Minnesota Digital Conservancy**

<http://conservancy.umn.edu/pol-preservation.jsp>

This document from the Office of Information Technology at the University of Minnesota outlines digital preservation support levels, file support, and preservation best practices for the University Digital Conservancy's preservation program. Released 2007.

### **University of Edinburgh: Managing Data**

<http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-support/research-data-mgmt>

This material by the University of Edinburgh's Information Services provides researchers with a range of information, from defining research data and funders policies and guidelines, to data management planning and methods for sharing research data. Released February 2010.

### **UK Data Archive Preservation Policy**

[http://www.data-archive.ac.uk/media/54776/ukda062-dps\\_preservationpolicy.pdf](http://www.data-archive.ac.uk/media/54776/ukda062-dps_preservationpolicy.pdf)

This document outlines the principles that allow for the active preservation of digital resources for use and re-use. Released June 2010.

## Roles, Services, and Skills

The reports and resources listed below are useful assessments and examples of the roles that data librarians, data scientists, and others may have in research data curation. These resources also characterize the skills that data curation requires, and ways in which curation activities are embodied in services to scientists and the larger community. Library and Information schools, degree or certificate programs, are not included.

### **Addressing the Research Data Gap: A Review of Novel Services for Libraries**

[http://www.carl-abrc.ca/about/working\\_groups/pdf/library\\_roles-final.pdf](http://www.carl-abrc.ca/about/working_groups/pdf/library_roles-final.pdf)

This 14-page report published by the Canadian Association of Research Libraries reviews exemplars of services that libraries can provide to meet research data needs. Services discussed include awareness and advocacy, training and support, discovery and access, archiving and preservation, and virtual research environments. Released March 2010.

### **Agenda for Developing E-Science in Research Libraries. Final Report and Recommendations to the Scholarly Communication Steering Committee, the Public Policies Affecting Research Libraries Steering Committee, and the Research, Teaching, and Learning Steering Committee**

[http://www.arl.org/bm~doc/ARL\\_EScience\\_final.pdf](http://www.arl.org/bm~doc/ARL_EScience_final.pdf)

This 26-page report lists recommendations for Association of Research Libraries (ARL) and its membership of North American research libraries for capacity-building and services, advocacy, and policies related to support for e-science and data curation.. Released November 2007.

### **ARL e-Science Survey Resource Page**

<http://www.arl.org/rtl/eresearch/escien/esciensurvey/surveyresearch.shtml>

The survey resource page presents results of the 2009 survey of Association of Research Libraries (ARL) members' e-science support, along with extensive lists of links to member programs, outreach, planning materials, grant-funded projects, and other data curation-related information.

### **DCC -- Data Management Courses and Training**

<http://www.dcc.ac.uk/node/8975>

The Digital Curation Centre (DCC) provides an up-to-date listing of international institutions providing training and degrees in data-related positions, along with links to roles and core skills descriptions for Data Creator, Data Manager, and Data Librarian or Data Scientist.

### **Dealing with Data: Roles, Rights, Responsibilities and Relationships**

[http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing\\_with\\_data\\_report-final.pdf](http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf)

In this 65-page UKOLN report, Liz Lyon provides a snapshot of the relationships and roles of the various stakeholders responsible for managing data in the UK, and a set of 10 recommendations for action. Released June 2007.

### **Open Science at Web-Scale: Optimising Participation and Predictive Potential**

<http://www.jisc.ac.uk/publications/reports/2009/opensciencerppt.aspx>

This Joint Information Systems Committee (JISC) report provides an examination of open science, especially open notebook science, citizen science, and data-driven predictive science. Of particular note is the examination of data curation roles and skill requirements for libraries to support open science. Released November 2009.

### **Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs**

<http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx>

A follow-up to a 2007 report by Liz Lyon, this Joint Information Systems Committee (JISC) report examines roles and skills needed for data scientists and curators, and the supply of skills to the current sector in the UK. July 2008.

### **To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering**

<http://www.arl.org/pp/access/nsfworkshop.shtml>

This report to the National Science Foundation from an Association of Research Libraries

workshop examines the role of academic libraries in supporting science and engineering through digital data curation.

### **Transforming Research Libraries: E-Science**

<http://www.arl.org/rtl/eresearch/escien/index.shtml>

This section of the Association of Research Libraries web site contains reports, articles and events pertaining to the changing roles and training/skill requirements for librarians as they plan and implement e-science and research data curation support services.

## **Sustainability and Cost Projections**

Most papers and reports mention that sustainability and funding are issues in data curation, but do not necessarily provide numbers or detailed outlines. The reports and studies listed below, written between 2007 and 2010, provide actual costs and/or detailed explanations of the issues regarding sustainability and funding.

### **A Shared Research Data Service for the UK**

<http://www.hefce.ac.uk/finance/shared/feasibility/show.asp?id=18&cat=6>

This report by the London School of Economics and Political Science investigates the feasibility and costs of developing and maintaining a managed shared research data service across the UK. It addresses the need not just for storage capacity, but also for active management of the creation, selection, ingestion, storage, retrieval, and preservation of research data. Released December 2008.

### **Disk and Tape Storage Cost Models**

[http://users.sdsc.edu/~mcdonald/content/papers/dt\\_cost.pdf](http://users.sdsc.edu/~mcdonald/content/papers/dt_cost.pdf)

This paper, written by Richard L. Moore, Jim D'Aoust, Robert H. McDonald, and David Minor, describes current estimates and projected costs of both disk and tape storage based on operational experience at the San Diego Supercomputer Center. Costs include storage hardware, supporting servers and related infrastructure, hardware maintenance, software licenses, floor space, utilities and labor costs. The paper also includes a comparison with current web-based commercial storage services. Released 2007.

### **Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age**

[http://www.nap.edu/catalog.php?record\\_id=12615](http://www.nap.edu/catalog.php?record_id=12615)

This report from the Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age discusses all aspects of data integrity, accessibility, and stewardship, including the sustainability and cost issues associated with the long-term care of research data. Released 2009.

### **Identifying Benefits Arising from the Curation and Open Sharing of Research Data**

<http://ie-repository.jisc.ac.uk/279/>

This report from the UK Higher Education and Research Institutions includes costs-benefits associated with data sharing including a business plan framework. Released January 2009

### **Infrastructure Planning and Data Curation: A Comparative Study of International Approaches to Enabling the Sharing of Research Data**

[http://www.jisc.ac.uk/media/documents/programmes/preservation/national\\_data\\_sharing\\_report\\_final.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/national_data_sharing_report_final.pdf)

This international study reviewed numerous methods of data sharing and the associated costs. Released November 2008.

### **Keeping Research Data Safe (Phase 1)**

<http://www.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx>

Released May 2008.

### **Keeping Research Data Safe (Phase 2)**

<http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx>

Discussion of cost models and variables affecting the long-term preservation costs for research data in UK universities. Released April 2010.

### **Open Data for Global Science**

[www.spatial.maine.edu/icfs/Uhlir-SchroederPaper.pdf](http://www.spatial.maine.edu/icfs/Uhlir-SchroederPaper.pdf)

Paul F. Uhlir and Peter Schröder review the opportunities, challenges and responsibilities to the global science research community associated with establishing an open data policy. Released June 2007.

#### **Researchers' Use of Academic Libraries and Their Services**

<http://www.rin.ac.uk/researchers-use-libraries>

This study by the Research Information Network and the Consortium of Research Libraries provides a view of how researchers interact with academic libraries in the UK and on the development of services academic libraries may want to provide researchers. Released April 2007.

#### **Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information: Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access**

<http://brtf.sdsc.edu/publications.html>

A National Science Foundation-funded analysis of models for sustainable and economically viable digital preservation for the preservation of digital information. Released February 2010.

## **Directories of Data Repositories**

The following resources are directories of repositories, listing distinct data centers and repositories for scientific data sets.

#### **Public Data Sets on Amazon Web Services**

<http://aws.amazon.com/publicdatasets/>

Amazon Web Services provides a centralized place to download public domain and non-proprietary astronomy, biology, chemistry and climatology data sets.

#### **Oceanographic Data Repositories**

<http://www.bco-dmo.org/data>

Funded by the National Science Foundation, the Biological and Chemical Oceanography Data Management Office (BCO-DMO) provides access to several oceanographic data repositories created by the US Joint Global Ocean Flux Study and US Global Ocean Ecosystem Dynamic programs.

#### **Distributed Data Curation Center: Other Data Repositories**

<http://d2c2.lib.purdue.edu/OtherRepositories.php>

Managed by Purdue University Libraries, the Distributed Data Curation Center lists of more than 50 open data repositories from a range of science disciplines.

#### **Gene Expression Omnibus**

<http://www.ncbi.nlm.nih.gov/geo/>

The Gene Expression Omnibus (GEO) is an open data repository which provides access to microarray, next-generation sequencing, and other forms of functional genomic data submitted by the scientific community.

#### **Global Change Master Directory**

<http://gcmd.nasa.gov/KeywordSearch/Home.do>

The Global Change Master Directory, maintained by the Earth Sciences Directorate at the National Aeronautics and Space Administration (NASA), provides access to more than 25,000 earth and environmental science data sets, relevant to global change and Earth science research.

#### **MIT Data Management and Publishing: Sharing Your Data**

<http://libraries.mit.edu/guides/subjects/data-management/publishing.html>

The MIT Libraries' subject guide on data management and publishing includes a list of open data repositories spanning the disciplines of astronomy, atmospheric science, biology, chemistry, earth science, oceanography and space science.

#### **Open Access Directory: Data Repositories**

[http://oad.simmons.edu/oadwiki/Data\\_repositories](http://oad.simmons.edu/oadwiki/Data_repositories)

Launched in 2008 and hosted by the Graduate School of Library and Information Science at

Simmons College, the Open Access Directory is a wiki that lists links to over 50 open data repositories in the disciplines of archaeology, biology, chemistry, environmental sciences, geology, geosciences and geospatial data, marine sciences, medicine and physics, as well as multidisciplinary open data repositories.

## Metadata Standards

The application of metadata enables discovery, use and proper citation of research data. The resources listed below do not constitute a comprehensive catalog of all metadata schemas for scientific datasets, but instead serve as pointers to indexes, aggregations, and communities of practice, with an emphasis on the most relevant metadata schemas for datasets within this webliography's scope.

### **Dublin Core Metadata Initiative (DCMI) Science and Metadata Community**

<http://dublincore.org/groups/sam/>

The DCMI Science and Metadata Community is a forum for individuals and organizations to exchange information and knowledge about metadata describing scientific data. The Community focuses on metadata challenges specific to scientific data curation, and solutions that will benefit from the architecture and global reach of the Dublin Core Metadata Initiative.

### **Darwin Core**

<http://rs.tdwg.org/dwc/>

Darwin Core is a metadata standard intended to facilitate the reference and sharing of biological diversity datasets.

### **Directory Interchange Format (DIF) Writer's Guide**

<http://gcmd.nasa.gov/User/difguide/>

DIF is a descriptive and standardized format for exchanging information about Earth science data.

### **Seeing Standards: A Visualization of the Metadata Universe**

<http://www.dlib.indiana.edu/~jenrile/metadatamap>

This web site, intended to assist planners with the selection and implementation of metadata standards, organizes metadata schemas in an easy-to-understand visual snapshot with an accompanying descriptive glossary. Metadata standards for data sets, such as Ecological Metadata Language (EML), Discovery Interchange Format (DIF) and the Virtual Solar Observatory Data Model, are included in the documentation.

### **ISO 19115:2003 Geographic Information - Metadata**

[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=26020](http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020)

Developed by the geospatial community, ISO 19115:2003 is a standard used to describe digital geographic data such as maps, charts, and textual documents as well as non-geographic data.

### **Preservation Metadata: Implementation Strategies (PREMIS)**

<http://www.loc.gov/standards/premis/>

PREMIS, the name of an international working group sponsored by OCLC and the Research Libraries Group (RLG), produced a report called PREMIS Data Dictionary for Preservation Metadata which includes information about the application and use of preservation metadata.

### **Science Data Literacy Project (SDL): Metadata Standards**

[http://sdl.syr.edu/?page\\_id=32\\*](http://sdl.syr.edu/?page_id=32*)

The SDL project is funded as part of the National Science Foundation's (NSF) effort to advance scientific progress through computer and technical infrastructure. This site links to metadata standards for the disciplines of astronomy, biology, geography, ecology and oceanography.

## Software and Middleware

Data producers and curators will find the following software tools suitable for different stages of the data curation lifecycle (e.g., data creation, data sharing, and data preservation). Not included are technical protocols or domain- or discipline-specific tools. While this list is current as of publication, it is worth noting that developers regularly improve software; over time new tools may supersede those on this list.

### **ARCHER Project**

<http://archer.edu.au/>

Completed in October 2008 with funding from the Australian Commonwealth Department of Education, Science and Training, the Australian Research Enabling environment includes software tools for the collection and retention of large data sets with associated metadata. Additionally, the eResearch tools provide a collaborative environment for data set annotation and discussion, and support research publication and access. The ARCHER toolset is freely available under the terms of the GNU General Public License.

### **The Dataverse Network Project**

<http://thedata.org/>

The Dataverse Network Project presents two options to interested researchers: 1) social science researchers can create a *dataverse*, or data archival space, at Harvard's Institute for Quantitative Social Science (IQSS) to publish their research data; or 2) researchers can install the open source dataverse network software to host multiple dataverses. The second option is appropriate for large archives or other institutions which are able to host the software and offer multiple dataverses as a service to researchers at the organization. While the first option is specifically geared toward social scientists, the second option is appropriate for researchers producing data in any field, including the natural sciences.

### **DSpace**

<http://dspace.org/>

Developed through a collaboration between Hewlett-Packard and MIT, DSpace is an open source software package used by many types of organizations, particularly academic, to host open digital repositories. According to statistics from the Registry of Open Access Repositories (<http://roar.eprints.org/view/software/>), it is the most widely used open source repository software. In July 2010, the DuraSpace organization recommended that the DSpace and Fedora communities investigate strategies to allow DSpace software to run on top of the Fedora platform. Examples of data repositories using DSpace include Dryad (<http://datadryad.org/>) and Edinburgh DataShare (<http://datashare.is.ed.ac.uk/>).

### **EPrints**

<http://www.eprints.org/>

Developed at the University of Southampton's School of Electronics and Computer Science, EPrints is one of the leading academic open source digital repository systems. Currently, there are 337 repositories worldwide utilizing EPrints. An example of a research data repository running on EPrints is eCrystals (<http://ecrystals.chem.soton.ac.uk/>).

### **eSciDoc Project**

<https://www.escidoc.org/>

Built upon the Fedora framework, eSciDoc provides an "open source e-research environment." In addition to a repository infrastructure, it provides related services which can be combined to address various data-related curation needs, such as data publication and preservation. It is possible to create virtual research environments (VREs) with this system.

### **Fedora**

<http://fedora-commons.org/>

The Flexible Extensible Digital Object Repository Architecture, developed jointly by the University of Virginia Library and Cornell University's Digital Library Research Group, is an open source repository platform which can be used as a foundation for a variety of repository applications; it does not include a front end for object discovery or delivery. The eSciDoc and Islandora projects (annotated in this webliography) are both built upon Fedora. It is expected that the Biophysical Repositories in the Lab (BRIL) repository (<http://bril.cerch.kcl.ac.uk/>) will run on Fedora.

### **gCube Framework**

<http://www.gcube-system.org/>

The gCube framework is an open source platform for the implementation of VREs which provides a grid-based data repository, as well as data analysis services and tools for publishing and sharing research data. It is based on the D4Science e-Infrastructure (<http://www.d4science.eu/>) in Europe.

### **Islandora**

<http://islandora.ca/>

An open source framework based on Fedora and the open source content management system Drupal (<http://drupal.org/>), Islandora is a digital asset management system developed by the Robertson Library at the University of Prince Edward Island. The project consists of a suite of software initiatives, including customized virtual research environments.

### **Integrated Rule-Oriented Data System (iRODS)**

<https://www.irods.org/index.php/>

Developed by the creators of Storage Resource Broker (see below) and their collaborators, iRODS is data grid software which uses policies, known as *rules*, to manage shared digital collections. Making use of data grids, curators implement these rules to build shareable collections within a federated network of repositories. For example, rules can make it possible for research data generated at one location to be accessed, manipulated, and/or preserved at other locations. iRODS is open source under a BSD license.

### **myExperiment**

<http://www.myexperiment.org/>

This site allows researchers to upload and share scientific workflows for others to execute. Additionally, it provides open source software which allows researchers to run their own instance of the software. Launched in November 2007 by a joint team from the universities of Southampton and Manchester in the UK, the social web site contains over 900 workflows covering a broad range of disciplines.

### **Project Trident**

<http://research.microsoft.com/en-us/collaboration/tools/trident.aspx>

In addition to allowing researchers to compose scientific workflows, Microsoft Research's Trident pairs with social networking sites such as myExperiment (see above) to share workflows publicly. Data curators can implement the open source tool to create a collaboration space where research outputs may be analyzed and shared.

### **Storage Resource Broker (SRB)**

[http://www.sdsc.edu/srb/index.php/Main\\_Page](http://www.sdsc.edu/srb/index.php/Main_Page)

The Storage Resource Broker (the precursor to iRODS, above) is middleware which manages shared collections across distributed organizations. While still supported, there are tools under development to enable migration from SRB to iRODS.

## **Assessment Tools**

The following tools are applicable to organizations interested in assessing the state of their digital research data assets.

### **Assessing Institutional Digital Assets (AIDA)**

<http://aida.jiscinvolve.org/>

Funded by Joint Information Systems Committee (JISC), the AIDA Project provides a toolkit (2nd edition, 2009; <http://aida.jiscinvolve.org/wp/toolkit/>) to allow institutions to perform assessments of their capacity, state of readiness, and overall capability for the curation of digital assets. The target audience of the project is institutions of higher education in the UK.

### **Data Curation Profiles (DCP)**

<http://datacurationprofiles.org>

A collaboration between the Distributed Data Curation Center of the Purdue University Libraries and the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign, this project investigates the practices and attitudes of researchers regarding the data they produce via research. Investigators conducted individual interviews with a cross disciplinary selection of researchers and developed case studies, or *profiles*, of their findings. The development of a DCP tool is in progress.

### **Data Asset Framework (DAF)**

<http://www.data-audit.eu/>

Originally called the Data Audit Framework, the DAF provides an audit methodology

([http://www.data-audit.eu/DAF\\_Methodology.pdf](http://www.data-audit.eu/DAF_Methodology.pdf)) for institutions to determine what research data assets they hold, and how those assets are managed. Funded by Joint Information Systems Committee (JISC) and led by the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow in association with the Digital Curation Centre, the framework also includes an online tool and registry (<http://www.data-audit.eu/tool.html>).

## Associations and Organizations

Most library associations are developing programs and resources in digital research data curation, and should be monitored for updates. These groups provide support in the form of policy development, training, networking, and research.

### **American Society for Information Science and Technology (ASIS&T)**

<http://www.asist.org/>

This group sponsors annual meetings, workshops, and symposia, including the 2010 Research Data Access and Preservation Summit in Phoenix, Arizona. Special interest groups provide avenues for professional development, including the Bioinformatics, Digital Libraries, and Scientific and Technical Information groups.

### **Association of European Research Libraries (Ligue Des Bibliothèques Européennes De Recherche, LIBER)**

<http://www.libereurope.eu/>

A working group of this association is focusing on the topics of e-science, research data, and workflows for the 2010-2012 biennium.

### **Association of Research Libraries (ARL)**

<http://www.arl.org/>

ARL, which includes libraries from North America, is concerned with the changing roles of research libraries, including the curation of research data. The association provides links to member activities, surveys, policy guidelines, reports, outreach resources, and training.

### **Australian National Data Service (ANDS)**

<http://ands.org.au/>

ANDS is an organization for all higher education providers and publicly funded research organizations in Australia. It investigates and develops policies, guidelines, and examples of research data ownership and access, including curation policies and tools for collecting and managing data.

### **Australian Partnership for Sustainable Repositories (APSR)**

<http://www.apsr.edu.au/>

This Australian organization provides outreach and education, and funds collaborative research and development projects for digital collections, including digital research data.

### **Coalition for Networked Information (CNI)**

<http://www.cni.org/>

CNI, sponsored by Educause and the Association of Research Libraries, supports projects, meetings, and conferences to build systems, standards, practices, and capacity for networked information. Task force meeting presentations and links to sponsored programs are available on the site.

### **Commerce, Energy, NASA, Defense Information Managers Group (CENDI)**

<http://www.cendi.gov/index.html>

Fourteen U.S. Federal Agencies make up this working group. Its goal is to improve efficiency in the areas of scientific and technical information capabilities. Interest areas in CENDI include metadata, taxonomies, preservation, and virtual libraries, and the group has provided a workshop on managing scientific data for its members.

### **Committee on Data for Science and Technology (CODATA)**

<http://www.codata.org/>

This committee of the International Council for Science (ICSU) works to improve accessibility and quality of scientific data sets, through working groups, workshops, publications, and

conferences.

#### **DataCite**

<http://datacite.org/>

This international consortium works to facilitate access to scientific research data through data registries and the promotion of research data as citable materials in the scientific record.

#### **Digital Curation Centre (DCC)**

<http://www.dcc.ac.uk/>

Created and funded by the United Kingdom's JISC (Joint Information Systems Committee), this organization supports and funds a large number of projects in data curation research, policies, tools and systems for UK higher education institutions.

#### **Digital Preservation Coalition (DPC)**

<http://www.dpconline.org/>

This not-for-profit coalition of both private/commercial and public organizations and individuals of the UK provides support for the adoption of digital preservation policies and practices.

#### **Dublin Core Metadata Initiative (DCMI)**

<http://dublincore.org/>

This open membership organization is dedicated to the development of metadata standards that are vital to describing scientific data.

#### **International Association of Scientific and Technical University Libraries (IATUL)**

<http://www.iatul.org/>

This association sponsors an annual conference with opportunities to network with librarians from around the world. The 2010 meeting included a variety of presentations on digital data curation (<http://docs.lib.purdue.edu/iatul2010/>).

#### **International Council for Scientific and Technical Information (ICSTI)**

<http://www.icsti.org/>

An international organization which sponsors scientific and technical projects, such as the integration of data citation with text, which culminated in the DataCite consortium.

#### **Joint Information Systems Committee (JISC)**

<http://www.jisc.ac.uk/>

Funded by UK higher education funding bodies, JISC manages an extensive list of projects, programs, and services for information technology innovation. These include data services and collections, digital repositories, open technologies, standards, and infrastructure.

#### **Research Data Strategy Working Group**

<http://data-donnees.gc.ca/eng/index.html>

A component of Research Data Canada, this working group draws its membership from libraries, research institutions, agencies and individuals, and is focused on solutions for managing Canadian research data.

#### **Research Information Network (RIN)**

<http://www.rin.ac.uk/>

The Research Information Network is funded by UK higher education and national libraries. RIN does open science case studies, research on data centers, and develops principles and guidelines for data stewardship.

#### **UK Research Data Service (UKRDS)**

<http://www.ukrds.ac.uk/>

This joint project funded by Joint Information Systems Committee (JISC) and the Higher Education Funding Council for England (HEFCE) is developing a planning and costing model for a national shared digital research data service for UK higher education.

## **Open Access Journals**

Listed below are journals whose primary focus is digital curation. The reader may also wish to investigate journals of particular science research disciplines, not included here, which may discuss research data curation topics.

#### **DCC -- Curation and Preservation Related Journals**

<http://www.dcc.ac.uk/resources/curation-journals>

This page lists a number of open-access titles on data curation and preservation.

#### **Data Science Journal**

<http://www.jstage.jst.go.jp/browse/dsj/-char/en/>

A publication of the Committee on Data for Science and Technology (CODATA) of the International Council for Science, this open access, peer-reviewed journal covers open data, data capture, analysis and visualization, policies, database development, and a wide range of other data curation-related topics.

#### **Database: The Journal of Biological Databases and Curation**

<http://database.oxfordjournals.org/>

Published by Oxford University Press, this open access journal covers data curation topics in all areas of biology, and topics related to addressing specific biology research questions are encouraged. Emphasizes biocuration but not to the exclusion of other material.

#### **International Journal of Digital Curation**

<http://www.ijdc.net/index.php/ijdc>

Published by UKOLN twice yearly, this peer-reviewed, open access electronic journal focuses solely on the curation of digital objects, including research data.

#### **International Journal of Spatial Data Infrastructures Research**

<http://ijsdir.jrc.ec.europa.eu/>

This peer-reviewed journal in support of spatial data, products, and underlying services and infrastructure is published by the Joint Research Centre of the European Commission.

## **E-mail Lists**

Many of the associations and organizations listed above have e-mail lists for members to share and discuss digital curation topics. The following are considered primary lists for digital data curation.

#### **Digital Curation Centre (DCC) Associates**

<http://www.dcc.ac.uk/community/dcc-associates>

This list provides a regular newsletter and announcements of DCC activities in digital data curation, which are frequently relevant to researchers and practitioners beyond the UK.

#### **Dublin Core Metadata Initiative (DCMI) Science and Metadata Community**

<http://dublincore.org/groups/sam/>

This list disseminates announcements about DCMI activities and other information for those interested in or working with metadata for scientific data.

#### **Digital Curation Google Group**

<http://groups.google.com/group/digital-curation>

This group provides a forum for discussions and information sharing about digital repository software, digital formats, technology, standards and interoperability, and other digital curation topics.

#### **Engineering and Science Informatics (ESI)**

<http://mailman.mit.edu/mailman/listinfo/esi>

The ESI list focuses on issues surrounding engineering and scientific data.

#### **Research Data Management List**

<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=RESEARCH-DATAMAN>

This list has some overlap with the DCC Associates list, but focuses on scientific research data, and incorporates discussions and announcements of activities in the UK and internationally.

## References

**Borgman, C., Wallis, J., Enyedy, N.** 2007. Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries* [Internet] [Cited November 17, 2010]. 7:17-30. Available from: <http://dx.doi.org/10.1007/s00799-007-0022-9>

**Choudhury, S.** 2009. Rethinking Scholarly Communication: Building Data Curation Infrastructure. [Internet] [Cited November 17, 2010]. Available from: <http://www.it.utah.edu/leadership/research/ciday/2009/notes/choudhury.pdf>

**Gold, A.** 2010. Data Curation and Libraries: Short-Term Developments, Long-Term Prospects. Office of the Dean (Library) [Internet]. [Cited November 17, 2010]. Available from: [http://digitalcommons.calpoly.edu/lib\\_dean/27](http://digitalcommons.calpoly.edu/lib_dean/27)

**Rusbridge C., Burnhill, P., Ross, S., Buneman, P., Giaretta, D., Lyon, L., Atkinson, M.** 2005. The Digital Curation Centre: A Vision for Digital Curation. In: *Proceedings From Local to Global: Data Interoperability--Challenges and Technologies*. Forte Village Resort, Sardinia, Italy. pp. 1-11. [Internet]. [Cited November 17, 2010]. Available from: <http://eprints.erpanet.org/82/>

**Wright, M., Sumner, T., Moore, R., and Koch, T.** 2007. Connecting digital libraries to eScience: the future of scientific scholarship. *International Journal of Digital Libraries*. [Internet] [Cited November 17, 2010]. 7:1-4. Available from: <http://www.springerlink.com/content/832616v17076317m/>