CODATA Task Group on data citation standards and best practices: Research and analysis results

As part of the activities of the CODATA Data Citation Standards and Practices Task Group, since its inception in 2010, we conducted an inventory of existing literature on data citation best practices and attribution activities. This document is the result of the collection of bibliographic sources, subsequent research, and corresponding analysis.

Prior versions of this bibliography have been published on the Task Group's website. [http://www.codata.org/taskgroups/TGdatacitation/Bilbliography_Links.html]. This version integrates all additional references gathered after the publication of the Summary Report of our August 2011 Workshop (sponsored jointly with the Board on Research Data and Information of the National Academy of Sciences), National Research Council. *For Attribution -- Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*. Washington, DC: The National Academies Press, 2012.

The corpus of this bibliography was created by Task Group members and consultants who performed online database searches and captured information sources that are directly or peripherally focused on data citation practices and attribution. These contributions were then shared via email or via Zotero, a reference management software tool. Many new entries were contributed over time by authors and reviewers of the Task Group's reports, interviewees from various stakeholder communities, and other interested persons.

As of this version, we have found 441 resources and organized them into 16 different formats that cover the many facets of citation, such as policies, infrastructure, research practices, and best practices development. We concentrated our efforts on sources that were published during the past five years, with the occasional older, seminal item included because of additional context and background. Each source contains links, notes, or abstracts where applicable or possible. Research papers comprise the bulk of the bibliography and were further classified into papers that comprise government and committee reports, and surveys and studies. The table below breaks down the total number of sources into types of formats and number of resources per format.

	Total Citations	Percentage of Citations
Blogs, Wikis, Web groups	30	6.80%
Books	13	2.95%
Citation Guides	31	7.03%
Citation Software & Repositories	44	9.98%
Conferences, Workshops, Symposia, Meetings	11	2.49%
Data Management Policies/Practices	4	0.91%
Journal issues devoted to data	3	0.68%
Op-eds, Newsletters, Press Releases, Memorandums	20	4.54%
Organizations, Committees	23	5.22%
Papers	137	31.07%
Papers (Reports): Government, Committee	27	6.12%
Papers: Surveys, Studies	31	7.03%
Posters, Charts	5	1.13%
Presentations, PPTs, Videos	16	3.63%
Standards	6	1.36%
Websites	40	9.07%
Grand Total	**441**	**100.00%**

Table 1. Total number of sources by formant and resource number per format.

Topics frequently covered in the literature included:

- Linked data, dynamic data, open data
- Data set management practices (general or for different scientific fields such as biology)
- Technology such as infrastructure & system architecture, unique identifiers, semantic web
- Digital data collection, attribution, contributor identifier, dissemination, collaboration and sharing, preservation, archival, verification, provenance
- The use of ontologies, repositories
- Data usage & metrics
- Data publishing
- Geospatial data management
- Citation practices & standards, metadata, policy & partnerships

While collecting this inventory, we found few policy standards applicable to digital data citations or consensus practices for data attribution. At most, we found scattered best practices, when available, varied among disciplines. Furthermore, many of the current data citation practices that we found reflected the traditional print data citation methods, but often did not fully address the requirements of 21st Century scientific digital data.

From our review of the literature we found a number of citation guidelines, some of which provide exemplars of consensus practices or best practices for a particular organization. We examined core elements across citation guidelines in our bibliography and created a chart representing our findings in Table 1 below. We found a wide range of practices regarding required elements in citation practices.

Citation Elements	**Altman & King**	**ANSI/ NISO Z39.29**	**DataCite / ANDS**	**Data-PASS**	**Data-verse**	**DCC Sage**	**Dryad**	**ESDS**	**FGDC-STD-001-1998 / IPIY**	**GBIF**	**GESIS**	**ICPSR**	**ISO 690**	**Mooney & Witt**	**OECD**	**PAN-GEA**	**PDS**	**Pub-lishers**	**SEDAC**
Author	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Title	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Date	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Publisher	x	x	x		x	x	x	x	x	x			x	x			x	x	x
Pub. Location									x										
Location		x				x							x						
Funder																			
Material Designator	x	x	x		x						x		x	x	x				x
Number of Records										x								x	
Edition	x		x		x	x									x			x	x
Handle				x															
UNF	x				x	x								x					
URL	x	x			x	x		x		x		x	x	x	x				x
URN				x															
DOI	x	x	x	x	x		x	x			x	x	x	x	x	x		x	
Accessed Date		x							x				x		x			x	x
Parent													x		x				
Version			x	x		x					x	x							
Editor									x										
Distributor									x		x							x	
Distr. Medium									x									x	

Table 2. Core elements across citation guides.

Our research yielded much information related to researcher's practices and approaches to data management and its use and reuse. The information generated on this topic and its focus indicates that the community is moving towards

addressing issues in regard to assisting researchers with data management. Below are the total sources, separated by format.

Blogs, Wikis, Web Groups

1. Altman, M. Blog (2012). Micah Altman's blog. Retrieved August 1, 2013 from the WWW: http://drmaltman.wordpress.com/
2. Bibliographic ontology specification group (n.d.). Retrieved August 1, 2013 from the WWW: http://groups.google.com/group/bibliographic-ontology-specification-group/about?hl=en

   Description: The Bibliographic Ontology provides main concepts and properties for describing citations and bibliographic references (i.e. quotes, books, articles, etc) on the Semantic Web. This is the mailing list for developers of the BIBO, and its tools and technologies.
3. Boettiger, C. (2012). Citing lab notebook entries. [Web log post]. Retrieved July 30, 2013 from the WWW: http://www.carlboettiger.info/2012/11/23/citing-lab-notebook-entries.html
4. Callaghan, S. (2012). Citing bytes – Adventures in data citation. Retrieved August 1, 2013 from the WWW: http://citingbytes.blogspot.com/2011/12/idcc-2011-notes-from-day-1-plenary.html
5. DataCite (2011, January 15). Tracking data citation [Web log post] Retrieved August 1, 2013 from the WWW: http://datacite.wordpress.com/2011/01/15/tracking-data-citation/
6. DataCite users Google group (n.d.). Retrieved August 1, 2013 from the WWW: https://groups.google.com/forum/#!forum/datacite-users
7. Data Pub (California Digital Library). Retrieved August 1, 2013 from the WWW: http://datapub.cdlib.org/
8. De Roure, D. (2010). Replacing the paper: The twelve rs of the e-research record." R&D Information Services. Retrieved August 1, 2013 from the WWW: http://blogs.nature.com/eresearch/2010/11/27/replacing-the-paper-the-twelve-rs-of-the-e-research-record
9. Digital preservation matters (2011). Retrieved August 1, 2013 from the WWW: http://preservationmatters.blogspot.com/2011/10/cite-datasets-and-link-to-publications.html
10. Dryad (2013). Data citation guidelines. Retrieved from https://www.datadryad.org/wiki/Citing_Data
11. Earth science information partner federation (ESIP). (2012) Data stewardship/citations. Retrieved from http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations
12. Figshare. Figshare partners with Open Access mega journal publisher PLOS. (2013 Jan 30) [Web log post]. Retrieved July 30, 2013 from the WWW: http://figshare.com/blog/figshare_partners_with_Open_Access_mega_journal_publisher_PLOS/68
13. Gipp, B. (2010). JabRef + automatic metadata extraction from PDF files. Retrieved August 1, 2013 from the WWW: http://gipp.com/jabref-automatic-metadata-extraction-from-pdf-files-like-mendeley-2
14. Gobbledygook. (Plos, Martin Fenner). Interview with Geoffrey Bilder (CrossRef). Retrieved August 1, 2013 from the WWW: http://blogs.plos.org/mfenner/2009/02/17/interview_with_geoffrey_bilder/
15. IASIST SIGDC (special interest group on data citation) (n.d.). Retrieved August 1, 2013 from the WWW:http://groups.google.com/group/iassist-sigdc/browse_thread/thread/abc7c7b28e0df580

    Description: Promotes awareness of data-related research and scholarship through data citation. Includes style guides from Mooney and Witt's poster session.
16. IDMB blog. (n.d.). Retrieved August 1, 2013 from the WWW: http://www.southamptondata.org/idmb-blog.html
17. IPAW wiki (2010). Retrieved August 1, 2013 from the WWW: http://tw.rpi.edu/portal/Main_Page
18. Interagency Working Group on Digital Data. (2009). Harnessing the power of digital data for science and society. Retrieved July 20, 2013 from the IT Law wiki http://itlaw.wikia.com/wiki/Interagency_Working_Group_on_Digital_Data

19. Knowledge blog. (n.d.). Retrieved August 1, 2013 from the WWW: http://knowledgeblog.org/

20. OJIMS. (n.d.). Retrieved August 1, 2013 from the WWW: http://proj.badc.rl.ac.uk/ojims

21. Pickrell, J. (2011, July 13). Why publish science in peer-reviewed journals? Genomes Unzipped. Retrieved August 1, 2013 from the WWW: http://www.genomesunzipped.org/2011/07/why-publish-science-in-peer-reviewed-journals.php

22. Piwowar, H. (2011). Research remix. Resources on data citation principles. Retrieved August 1, 2013 from the WWW: http://researchremix.wordpress.com/2011/05/17/resources-on-data-citation-principles

23. Plag, Hans-Peter. (n.d.). Retrieved form http://www.gstss.org/blogs/data_citation_blog.php

24. RDA: Case statement: Data citation – Making research data citable. (n.d). Retrieved August 1, 2013 from the WWW: https://rd-alliance.org/case-statement-data-citation-making-research-data-citable.html

25. SageCite blog. (n.d.). Retrieved August 1, 2013 from the WWW: http://blogs.ukoln.ac.uk/sagecite/

    Description: Produced a demonstrator citation service for network models, workflows and associated data in the Sage Commons, using a linked data approach.

26. Saller, C. (2011). Citation obsession? Dream on. *The Chronicle of Higher Education*. Retrieved August 1, 2013 from the WWW: http://chronicle.com/blogs/linguafranca/2011/11/03/citation-obsession-dream-on/

27. Shotton, David. Semantic publishing. Retrieved August 1, 2013 from the WWW:http://semanticpublishing.wordpress.com/2011/06/30/how-to-cite-data//

28. The Signal: Digital Preservation (LC blog, n.d.). Link persistence, website persistence. Retrieved August 1, 2013 from the WWW: http://blogs.loc.gov/digitalpreservation/2013/04/link-persistence-website-persistence/

29. Technology Watch Report (TWR, n.d.): Standards in metadata. Retrieved August 1, 2013 from the WWW: http://metadaten-twr.org

30. W3C provenance working group standardization activity (2013). Retrieved August 1, 2013 from the WWW: http://www.w3.org/2011/prov/wiki/Main_Page

Books

1. Altman, M., Gill, J., & McDonald, M. (2003). *Numerical issues in statistical computing for the social scientist*. New York, NY: John Wiley & Sons

   Description: Provides readers with a unique practical guidebook to the numerical methods underlying computerized statistical calculations specific to these fields. Highlights include: a focus on problems occurring in maximum likelihood IStata, LIMDEP, SPSS, WinBUGS, and MATLAB®); a guide to choosing accurate statistical packages; discussions of a multitude of computationally intensive statistical approaches such as ecological inference, Markov chain Monte Carlo, and spatial regression analysis; emphasis on specific numerical problems, statistical procedures, and their applications in the field; replications and re-analysis of published social science research, using innovative numerical methods; key numerical estimation issues along with the means of avoiding common pitfalls; a related Web site includes test data for use in demonstrating numerical problems; code for applying the original methods described in the book, and an online bibliography of Web resources for the statistical computation.

2. Berns, K. I., Bond, E. C., & Manning, F. J. (1996). *Resource sharing in biomedical research*. Washington, D.C.: National Academy Press. Retrieved August 1, 2013 from the WWW: http://www.nap.edu/catalog.php?record_id=5429

   Description: The book identifies six diverse case studies that share material or equipment with the scientific community at large. Common strength and problems are identified in the six cases, as well as a series of recommendations aimed at sharing in biomedical research.

3. Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet.* Cambridge, MA: The MIT Press

Description: Explores the technical, social, legal, and economic aspects of the kind of infrastructure that we should be building for scholarly research in the twenty-first century. Borgman describes the roles that information technology plays at every stage in the life cycle of a research project and contrasts these new capabilities with the relatively stable system of scholarly communication, which remains based on publishing in journals, books, and conference proceedings. No framework for the impending "data deluge" exists comparable to that for publishing. Analyzing scholarly practices in the sciences, social sciences, and humanities, Borgman compares each discipline's approach to infrastructure issues. In the process, she challenges the many stakeholders in the scholarly infrastructure—scholars, publishers, libraries, funding agencies, and others—to look beyond their own domains to address the interaction of technical, legal, economic, social, political, and disciplinary concerns.

4. Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, and National Academy of Sciences. (2009). *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. Washington, D.C.: National Academies Press. Retrieved August 1, 2013 from the WWW: http://www.nap.edu/catalog.php?record_id=12615

5. Davidson, S. B., & Freire, J. (2008). Provenance and scientific workflows: challenges and opportunities. *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. Retrieved July 30, 2013 from the WWW: http://bigdata.poly.edu/~juliana/pub/freire-tutorial-sigmod2008.pdf

6. Fetterer, F. (2009). Data management best practices for sea ice observation. In H. Eicken (Ed.). *Field Techniques for Sea-Ice Research* (pp. 395-404). Fairbanks, AK: University of Alaska Press. Retrieved August 1, 2013 from the WWW: http://nsidc.org/about/bios/fetterer.html

   Description: The first comprehensive research done on sea-ice field techniques, this volume will be indispensable for the study of northern sea ice and a must-have for scientists in the field of climate change research.

7. Geoscience Information Society, European Association of Science Editors (1999). *Science editing and information management: Proceedings of the second international Aese/ Cbe /Ease joint meeting.* Out of Print; limited availability. Retrieved August 1, 2013 from the WWW: http://www.amazon.com/dp/0934485305

8. Heath, T., & Bizer, C. (2011). *Linked data: evolving the web into a global data space*. San Rafael, CA: Morgan & Claypool

   Description: We provide readers with a detailed technical introduction to Linked Data. We begin by outlining the basic principles of Linked Data, including coverage of relevant aspects of Web architecture. The remainder of the text is based around two main themes - the publication and consumption of Linked Data. Drawing on a practical Linked Data scenario, we provide guidance and best practices on: architectural approaches to publishing Linked Data; choosing URIs and vocabularies to identify and describe resources; deciding what data to return in a description of a resource on the Web; methods and frameworks for automated linking of data sets; and testing and debugging approaches for Linked Data deployments. We give an overview of existing Linked Data applications and then examine the architectures that are used to consume Linked Data from the Web, alongside existing tools and frameworks that enable these.

9. Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). *The fourth paradigm: Data intensive scientific discovery*. Redmond, WA: Microsoft Research. Retrieved July 30, 2013 from the WWW: http://research.microsoft.com/en-us/collaboration/fourthparadigm/

10. Manson, C. J. (1999). *Science editing and information management: Proceedings of the Second International AESE/CBE/EASE Joint Meeting, Part Three*. Alexandria, VA: Geoscience Information Society

11. Novak, K., Altman, M., Broch, E., Carroll, J. M., Clemins, P. J., Fournier, D., Laevart, C., et al. (2011). *Communicating science and engineering data in the information age*. Washington, D.C.: The National Academies Press. Retrieved August 1, 2013 from the WWW: http://www.nap.edu/catalog.php?record_id=13282

Description: Includes recommendations to improve NCSES's dissemination program and improve data user engagement. This report includes recommendations such as NCSES's transition to a dissemination framework that emphasizes database management rather than data presentation, and that NCSES analyze the results of its initial online consumer survey and refine it over time. The implementation of the report's recommendations should be undertaken within an overall framework that accords priority to the basic quality of the data and the fundamentals of dissemination, then to significant enhancements that are achievable in the short term, while laying the groundwork for other long-term improvements.

12. Pryor, G. (Ed.). (2012). *Managing research data*. Oxon, UK: Facet Publishing

    Description: This edited collection, bringing together leading figures in the field from the UK and around the world, provides an introduction to all the key data issues facing the HE and information management communities.

13. Uhlir, P.F., (Ed.) (2012). *For attribution: Developing data attribution and citation practices and standards: Summary of an international workshop.* Washington, D.C: National Academies Press. Retrieved July 30, 2013 from the WWW: http://www.nap.edu/catalog.php?record_id=13564

Citation Guides

Library Resource Guides on Data Citation

1. Cambridge. Retrieved August 1, 2013 from the WWW: http://www.lib.cam.ac.uk/dataman/pages/citations.html
2. Minnesota. Retrieved August 1, 2013 from the WWW: http://www.lib.umn.edu/datamanagement/cite
3. MIT. Retrieved August 1, 2013 from the WWW: http://libraries.mit.edu/guides/subjects/data/access/citing.html
4. MSU. Retrieved August 1, 2013 from the WWW: http://libguides.lib.msu.edu/citedata
5. Oregon. Retrieved August 1, 2013 from the WWW: http://libweb.uoregon.edu/datamanagement/citingdata.html
6. Purdue. Retrieved August 1, 2013 from the WWW: http://guides.lib.purdue.edu/datacitation
7. Toronto. Retrieved August 1, 2013 from the WWW: http://datalib.chass.utoronto.ca/caq/citation.doc
8. UCambridge. Retrieved August 1, 2013 from the WWW: http://www.lib.cam.ac.uk/dataman/pages/citations.html
9. UMinn. Retrieved August 1, 2013 from the WWW: http://www.lib.umn.edu/datamanagement/cite
10. UVirginia. Retrieved August 1, 2013 from the WWW: http://dmconsult.library.virginia.edu/plan/citing-data/
11. UWM. Retrieved August 1, 2013 from the WWW: http://www4.uwm.edu/libraries/AGSL/agsgis/find.cfm

Non-Library Guides to Data Citation

1. ANDS. Data citation. Retrieved August 1, 2013 from the WWW: http://www.ands.org.au/guides/data-citation-awareness.pdf
2. Argonne National Laboratory. Argonne premium coal samples citation form. Retrieved August 1, 2013 from the WWW: http://web.anl.gov/PCS/citation.html
3. Ball, A., & Duke, M. (2012). How to cite data sets and link to publications. *DCC how-to guides.* Edinburgh: Digital Curation Centre. Retrieved July 30, 2013 from the WWW: http://www.dcc.ac.uk/resources/how-guides

    Description: Provides a working knowledge of the issues, challenges, and solutions to problems such as granularity, microattribution, contributor identifiers (ORCHID, ISNI), and placement of data citations. Also discusses citation infrastructures such as citation notification service (CLADDIER), Nano publications, Citation Typing Ontology, repositories, and implementation issues including manual and

automatic use of citations and dynamic datasets. This guide should interest researchers and principal investigators working on data-led research, as well as the data repositories with which they work.

4. DataCite. Retrieved August 1, 2013 from the WWW: http://www.datacite.org/whycitedata
5. Dryad. Retrieved August 1, 2013 from the WWW: http://www.datadryad.org/using
6. EOL. Retrieved August 1, 2013 from the WWW: http://eol.org/info/citing
7. Federal Geographic Data Committee. (2003). Managing historical geospatial data records. Retrieved August 1, 2013 from the WWW: http://www.fgdc.gov/library/factsheets/documents/histdata.pdf

   Description: The development of a National Spatial Data Infrastructure (NSDI) is an important step in ensuring the Nation's Economic, environmental and scientific well-being. The National Archives and Records Administration (NARA) is the Federal agency responsible for acquiring, preserving, and making available those records of enduring value created or received by various components of the Federal Government.
8. GESIS Data Archive. Retrieved August 1, 2013 from the WWW: http://www.gesis.org/en/services/data-analysis/data-archive-service/citation-of-research-data/
9. GEOSS Data Citation. Retrieved August 1, 2013 from the WWW: http://www.gstss.org/library/GEOSS_Data_Citation_Guidelines_V2.0.pdf
10. ICPSR. Retrieved August 1, 2013 from the WWW: http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp
11. NAS. Retrieved August 1, 2013 from the WWW: http://history.nasa.gov/citeguide.html
12. NASA PDS. Retrieved August 1, 2013 from the WWW: http://ppi.pds.nasa.gov/citations_policy.jsp
13. NOAA. Retrieved August 1, 2013 from the WWW: http://www.ncdc.noaa.gov/paleo/citation.html
14. Pensoft. Retrieved August 1, 2013 from the WWW: http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf
15. SAEON. Retrieved August 1, 2013 from the WWW: http://data.saeon.ac.za/management/it-governance/G373.2%20Policy%20and%20Guidelines/G373.2.1%20Guidelines%20and%20Best%20Practice/
16. Socioeconomic Data and Applications Center (SEDAC). Retrieved August 1, 2013 from the WWW: http://sedac.ciesin.org/citations
17. Statistic Canada. Retrieved August 1, 2013 from the WWW: http://www.statcan.gc.ca/pub/12-591-x/2009001/steps-etapes-eng.htm
18. UK Data Archive. Retrieved August 1, 2013 from the WWW: http://data-archive.ac.uk/conditions/citing-data
19. United States Department of Agriculture (2012). Soil data access- citation. Retrieved August 1, 2013 from the WWW: http://sdmdataaccess.nrcs.usda.gov/Citation.htm
20. USGS LP DAAC. Retrieved August 1, 2013 from the WWW: https://lpdaac.usgs.gov/about/citing_lp_daac_and_data

Citation Software and Repositories

1. ArXiv (Cornell). Retrieved August 1, 2013 from the WWW: http://arxiv.org/
2. Australian National Data Service. Retrieved August 1, 2013 from the WWW: http://www.ands.org.au/
3. Australian Research Collaborative Services. Retrieved August 1, 2013 from the WWW: http://le.unimelb.edu.au/research/arcs.html
4. BGI (Beijing Genomics Institute) Cloud Computing. Retrieved August 1, 2013 from the WWW: http://www.genomics.cn/index

5. BMC_BL_Data_repositories (list). Retrieved August 1, 2013 from the WWW: https://docs.google.com/spreadsheet/ccc?authkey=COmDvOUB&key=0Aok0Od_Hhd1XdEdiRXVCbDlFWk8wNW5FYlBBTndyaVE&hl=en_US&authkey=COmDvOUB#gid=0

   Description: Lists 155 domain-specific and general data repositories. Includes name, website, subject area, funding model, restrictions, license agreement, county, identifiers, abbreviation, notes, representatives, and standards

6. Cambridge Crystallographic Data Centre. Retrieved August 1, 2013 from the WWW: http://www.ccdc.cam.ac.uk/
7. Data.gov. Retrieved August 1, 2013 from the WWW: www.data.gov
8. DATAPASS. Retrieved August 1, 2013 from the WWW: http://www.icpsr.umich.edu/icpsrweb/content/DATAPASS/citations.html
9. DanBIF. Retrieved August 1, 2013 from the WWW: http://www.danbif.dk/
10. Dataverse. Retrieved August 1, 2013 from the WWW: http://thedata.org/
11. DTOL. Retrieved August 1, 2013 from the WWW: https://sites.google.com/site/datatolproject/schema
12. Dryad. Retrieved August 1, 2013 from the WWW: http://datadryad.org/

    Description: Established a UK mirror of the Dryad data repository, extended its support to new publishers and disciplines, and developed a sustainability plan and performance metrics.

13. dSPACE. Retrieved August 1, 2013 from the WWW: http://www.dspace.com/en/inc/home.cfm
14. EBI. Retrieved August 1, 2013 from the WWW: http://www.ebi.ac.uk/
15. ESDS. Retrieved August 1, 2013 from the WWW: http://www.esds.ac.uk/international/.

    Video here: Retrieved August 1, 2013 from the WWW: http://www.youtube.com/watch?v=NDrNHRjtd4g

16. EndNote. Retrieved August 1, 2013 from the WWW: http://www.endnote.com
17. Figshare. Retrieved August 1, 2013 from the WWW: http://figshare.com/
18. FISH.Link. Retrieved August 1, 2013 from the WWW: http://www.jisc.ac.uk/whatwedo/programmes/mrd/clip/fishlink.aspx

    Description: Produced tools for converting and mapping freshwater biology data to linked data, while supporting semantic markup, attribution and provenance

19. Galaxy. Retrieved August 1, 2013 from the WWW: http://galaxy.psu.edu/
20. GBIF. Retrieved August 1, 2013 from the WWW: http://www.gbif.org/
21. GenBank. Retrieved August 1, 2013 from the WWW: http://www.ncbi.nlm.nih.gov/genbank/
22. Giga Science (& British Library). Retrieved August 1, 2013 from the WWW: http://www.gigasciencejournal.com/
23. ICPSR. Retrieved August 1, 2013 from the WWW: http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp
24. INSPIRE SDI. Retrieved August 1, 2013 from the WWW: http://www.intergraph.com/learnmore/sgi/government/inspire.aspx

    Long term preservation of data here (PPT): http://inspire.jrc.ec.europa.eu/events/conferences/inspire_2010/presentations/55_pdf_presentation.pdf

25. International Virtual Observatory Alliance. Retrieved August 1, 2013 from the WWW: http://www.ivoa.net/
26. LOCKSS. Retrieved August 1, 2013 from the WWW: http://www.lockss.org/
27. Mendeley. Retrieved August 1, 2013 from the WWW: http://www.mendeley.com

28. Mint (Molecular INTeraction Database). Retrieved August 1, 2013 from the WWW: http://160.80.34.4/mint/Welcome.do

29. National Snow and Ice Data Center (NSIDC). Retrieved August 1, 2013 from the WWW: http://nsidc.org/

30. NERC. Retrieved August 1, 2013 from the WWW: http://www.nerc.ac.uk/

31. NGDA. Retrieved August 1, 2013 from the WWW: http://www.ngda.org/

32. ORCID. Retrieved August 1, 2013 from the WWW: http://about.orcid.org/

33. ORNL DAAC. Retrieved August 1, 2013 from the WWW: http://daac.ornl.gov/

34. PANGAEA. Retrieved August 1, 2013 from the WWW: http://www.pangaea.de/

35. Polar Information Commons. Retrieved August 1, 2013 from the WWW: http://www.polarcommons.org/ethics-and-norms-of-data-sharing.php

36. PDB (Protein Data Bank). Retrieved August 1, 2013 from the WWW: http://www.rcsb.org/pdb/home/home.do

37. RefWorks. Retrieved August 1, 2013 from the WWW: http://www.refworks.com

38. SAEON. Retrieved August 1, 2013 from the WWW: http://www.saeon.ac.za/

39. SAGECite. Retrieved August 1, 2013 from the WWW: http://www.sagebase.org/.

    Description: Produced a demonstrator citation service for network models, workflows and associated data in the Sage Commons, using a linked data approach.

40. SEAD (Sustainable Environment Actionable Data). Retrieved August 1, 2013 from the WWW: http://sead-data.net/

41. SND. Retrieved August 1, 2013 from the WWW: http://snd.gu.se/en

42. UK Data Service. Retrieved August 1, 2013 from the WWW: http://ukdataservice.ac.uk/

43. UnitProt (Universal Protein Resource Knowledgebase). Retrieved August 1, 2013 from the WWW: http://www.uniprot.org/

44. Zotero. Retrieved August 1, 2013 from the WWW: http://www.zotero.org

Conferences, Workshops, Symposia, Meetings

1. Beyond the PDF. (201, Jan). Workshop , San Diego, CA. Retrieved from Retrieved August 1, 2013 from the WWW: https://sites.google.com/site/beyondthepdf/

2. BRDI (2011, Aug). Developing data attribution and citation practices and standards. An international symposium and workshop, Berkeley, CA. Retrieved August 1, 2013 from the WWW: http://sites.nationalacademies.org/PGA/brdi/PGA_064019

3. Bridging data lifecycles: Tracking data use via data citations data workshop. (2012, Apr). University Corporation for Atmospheric Research (UCAR), Boulder, Co. Retrieved August 1, 2013 from the WWW: http://library.ucar.edu/data_workshop/

4. CLADDIER (2007, May). Linking data and publications in the environmental sciences: CLADDIER project workshop, Southampton, United Kingdom. Retrieved August 1, 2013 from the WWW: http://eprints.soton.ac.uk/46207/

5. European Science Foundation (2007, Sept). Shared responsibilities in sharing research data: Policies and partnerships. Workshop at the Berlin 5 Conference, Padua, Italy. Retrieved August 1, 2013 from the WWW: http://www.knowledge-exchange.info/Default.aspx?ID=66&M=News&PID=177&NewsID=24

6. Harvard University (2011, May). Principles of data citation workshop, sponsored by Quantitative Social Science, Harvard, MA. Retrieved August 1, 2013 from the WWW: http://projects.iq.harvard.edu/datacitation_workshop/

7. IASSIST data science professionals: A global community of sharing (2011, May-June). Conference, Vancouver, Canada. Retrieved August 1, 2013 from the WWW: http://www.iassistdata.org/conferences/archive/2011

8. Kelly, M.C. (2008, Oct). NISO thought leader meeting on research data. Baltimore, MD. Retrieved August 1, 2013 from the WWW: http://www.niso.org/topics/tl/NISOTLDataReportDraft.pdf

9. Metadata for managing scientific research data. (2012, Aug). Webinar. Retrieved August 1, 2013 from the WWW: http://www.niso.org/news/events/2012/dcmi/scientific_data/

10. Meeting with ocean science journal editors (2008, Dec). San Francisco, CA. Retrieved August 1, 2013 from the WWW: http://www.scor-int.org/Project_Summit_3/Data_Publication.pdf

11. Persistent identifiers for the social sciences (2011, Feb) Workshop sponsored by the IDSC of IZA/Gesis/RatSWD, Bonn, Germany. Retrieved August 1, 2013 from the WWW: http://www.iza.org/conference_files/PeIdSS2011/viewProgram?conf_id=2013

Data Management Policies/Practices

1. Availability of supporting data. (2013). BioMed Central. Retrieved August 1, 2013 from the WWW: http://www.biomedcentral.com/about/supportingdata

2. Engineering and Physical Sciences Research Council (2011). *ESPRC policy framework on research data*. Retrieved July 30, 2013 from the WWW: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx

3. Oxford policy on the management of research records. (2012). Retrieved August 1, 2013 from the WWW: http://www.admin.ox.ac.uk/rdm/managedata/policy/

4. Research data management policy. (2011). Retrieved August 1, 2013 from the WWW: http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy

Journal issues devoted to data

1. *The Economist*. (2010) Special report: Managing information. Retrieved August 1, 2013 from the WWW: http://www.economist.com/printedition/2010-02-27

2. *Nature, 455*(7209), 1-136. (2008). Special issue: Big data. Retrieved August 1, 2013 from the WWW: http://www.nature.com/nature/journal/v455/n7209/

3. *Science, 331*(6018), 639-806. (2011). Special issue: Dealing with data.

Op-eds, Newsletters, Press Releases, Memorandums

1. Berman, F. (2010). We need a research data census. *Communications of the ACM, 53*(12), 39-41. doi:10.1145/1859204.1859220

Description: The increasing volume of research data highlights the need for reliable, cost-effective data storage and preservation at the national scale.

2. Borowski, C. (2011). Enough is enough. *Journal of Experimental Medicine, 208*(7), 1337. Retrieved August 1, 2013 from the WWW: http://jem.rupress.org/content/early/2011/06/01/jem.20111061.full.pdf

Description: Complaints about the overabundance of supplementary information in primary research articles have increased in decibel and frequency in the past several years and are now at cacophonous levels. Reviewers and editors warn that they do not have time to scrutinize it. Authors contend that the effort and money needed to produce it exceeds that reasonably spent on a single publication. How often readers actually look at supplemental information is unclear, and most journal websites offer the supplement as an optional download.

3. Cook, R. (2008). Citations to published data sets. *Fluxletter, 1*(4), 4-5. Retrieved August 1, 2013 from the WWW: http://daac.ornl.gov/ornl_daac_citations_200812.pdf

4. Rooyen, H. (Council for Scientific and Industrial Research) (2010). *Data on southern oceans now freely available*. [Press Release]. Retrieved August 1, 2013 from the WWW: http://ntww1.csir.co.za/plsql/ptl0002/PTL0002_PGE157_MEDIA_REL?MEDIA_RELEASE_NO=7523509

   Description: The article gives a brief description of the different data types that are obtained and how the data is stored. The data is scanned before for errors and then it is saved or used by researchers. SADCO works in conjunction with the CSIR. They collect their data on the oceans from cruises, whose names are listed in an online inventory.

5. Credit Where Credit Is Due. [Editorial]. (2009, Dec 17). *Nature* 462: 825. doi:10.1038/462825

6. Datacite. (2009). Datacite memorandum of understanding. Retrieved August 1, 2013 from the WWW: http://datacite.org/datacite_memo_understanding

7. Data's shameful neglect. (10 Sept 2009). Nature [editorial]. Retrieved August 1, 2013 from the WWW: http://www.nature.com/nature/journal/v461/n7261/full/461145a.html

8. European Commission. (2011a). Digital agenda: Turning government data into gold. [Press release]. Retrieved July 30, 2013 from the WWW: http://europa.eu/rapid/press-release_IP-11-1524_en.htm

9. Geoscience Information Society. (2004). *Geoscience Information Society Newsletter*, 210. Retrieved August 1, 2013 from the WWW: http://www.geoinfo.org/GSIS_Newsletter/200410.pdf

10. Helly, J. (1998). New concepts of publication. *Nature*, *393*, 1998. Retrieved August 1, 2013 from the WWW: http://www.nature.com/nature/journal/v393/n6681/full/393107c0.html

11. Knecht, L, Auld V. A., & McGhee M. (2001). Changes in the treatment of chemical data in MEDLINE® citations. *National Library of Medicine Technical Bulletin*, 323. Retrieved August 1, 2013 from the WWW: http://www.nlm.nih.gov/pubs/techbull/nd01/nd01_mesh_chemical.html

12. Kolowich, S. (2011). Killing peer review. *Inside Higher Ed*. Retrieved August 1, 2013 from the WWW: http://www.insidehighered.com/news/2011/07/19/debate_over_whether_social_web_sites_can_replace_peer_review

13. Lowry, R., Urban, E., & Pissierssens, P. (2009). A new approach to data publication in ocean sciences. *Eos*, *90*(50), 484. doi:10.1029/2009EO500004

14. Maunsell, J. (2010). Announcement regarding supplemental material. *The Journal of Neuroscience*, *30*(32), 10599-10600. Retrieved August 1, 2013 from the WWW: http://www.jneurosci.org/content/30/32/10599.full

   Description: Beginning November 1, 2010, The Journal of Neuroscience will no longer allow authors to include supplemental material when they submit new manuscripts and will no longer host supplemental material on its web site for those articles.

15. NISO. (2005). NISO-sponsored INFO URI scheme gets thumbs up from IETF group. Retrieved August 1, 2013 from the WWW: http://www.niso.org/news/pr/view?item_key=4b8a9e2d84fe28e5559d725eb6acd6fd9b1eb53d

16. NISO Launches New Initiative to Develop Standard for Open Access Metadata and Indicators (2013, Feb 7). Retrieved August 1, 2013 from the WWW: http://www.niso.org/news/pr/view?item_key=d2e5f409bc6af6b7f504a10edf0329203ffec6f9

17. Paving the way to an open scientific information space: OpenAIREplus – linking peer-reviewed literature to associated data. [press release] (2011). Retrieved July 30, 2013 from the WWW: https://www.openaire.eu/en/component/content/article/76-highlights/326-openaireplus-press-release

18. PR Newswire. (2010). Elsevier and PANGAEA take next step in connecting research articles to data. *United Business Media*. Retrieved August 1, 2013 from the WWW: http://www.prnewswire.com/news-releases/elsevier-and-pangaea-take-next-step-in-connecting-research-articles-to-data-99533624.html

19. Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. Retrieved August 1, 2013 from the WWW: http://altmetrics.org/manifesto/

20. Request for Information (RFI): Input on Development of a NIH Data Catalog. Retrieved August 1, 2013 from the WWW: http://grants.nih.gov/grants/guide/notice-files/NOT-HG-13-011.html

Organizations, Committees

1. ANDS. Retrieved August 1, 2013 from the WWW: http://www.ands.org.au/
2. ASIS&T SIG/MET. Retrieved August 1, 2013 from the WWW: http://www.asis.org/SIG/met.html
3. CODATA Task Force on Data Citation. Retrieved August 1, 2013 from the WWW: http://www.codata.org/taskgroups/TGdatacitation/
4. Corporation for National Research Initiatives. Retrieved August 1, 2013 from the WWW: http://www.cnri.reston.va.us/about_cnri.html
5. CSIR. Retrieved August 1, 2013 from the WWW: http://www.csir.co.za/nre/ecosystems/Geoportal.html
6. DataCite. Retrieved August 1, 2013 from the WWW: http://datacite.org
7. Data.gov. Retrieved August 1, 2013 from the WWW: http://www.data.gov/
8. Dataverse. Retrieved August 1, 2013 from the WWW: http://thedata.org/
9. Digital Curation Centre (DCC). Retrieved August 1, 2013 from the WWW: http://www.dcc.ac.uk/
10. ESIP. Retrieved August 1, 2013 from the WWW: http://www.esipfed.org/
11. FGDC. Retrieved August 1, 2013 from the WWW: http://www.fgdc.gov/
12. International Polar Year. Retrieved August 1, 2013 from the WWW: http://classic.ipy.org/international/joint-committee/data-management.htm
13. JISC. Retrieved August 1, 2013 from the WWW: http://www.jisc.ac.uk/aboutus.aspx
14. NDIPP (Library of Congress). Retrieved August 1, 2013 from the WWW: http://www.digitalpreservation.gov/
15. OAI. Retrieved August 1, 2013 from the WWW: http://www.openarchives.org/
16. OSTI. Retrieved August 1, 2013 from the WWW: http://www.osti.gov/
17. PARSE Insight. Retrieved August 1, 2013 from the WWW: http://www.parse-insight.eu/project.php.

    Latest report here: Retrieved August 1, 2013 from the WWW: http://www.parse-insight.eu/downloads/PARSE-Insight_D2-2_Roadmap.pdf
18. Research Data Canada. National Consultation on Access to Scientific Research Data (NCASRD), Canada. Retrieved August 1, 2013 from the WWW: http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/ncasrd/
19. Southampton Data Management. Retrieved August 1, 2013 from the WWW: http://www.southamptondata.org/
20. SAEON. Retrieved August 1, 2013 from the WWW: http://www.saeon.ac.za/

    Data policy here: Retrieved August 1, 2013 from the WWW: http://data.saeon.ac.za/documentation/it-governance/policies-and-guidelines/data-policy-_stand-alone_.pdf/view
21. SURF (SURFShare). Retrieved August 1, 2013 from the WWW: http://www.surffoundation.nl/en/themas/openonderzoek/Pages/Default.aspx
22. TIB. Retrieved August 1, 2013 from the WWW: http://www.tib-hannover.de/en/
23. UKOLN. Retrieved August 1, 2013 from the WWW: http://www.ukoln.ac.uk/

Papers

1. Aalbersberg, I., & Kahler, O. (2011). Supporting science through the interoperability of data and articles, *D-Lib Magazine 17*(½). doi:10.1045/january2011-aalbersberg

Description: This article presents an overview of how Elsevier as a scientific publisher with over 2,000 journals gives context to articles that are available on their full-text platform SciVerse ScienceDirect, by linking out to externally hosted data at the article level, at the entity level, and in a deeply integrated way. With this overview, Elsevier invites dataset repositories to collaborate with publishers to create an optimal interoperability between the formal scientific literature and the associated research data — improving the scientific workflow and ultimately supporting science.

2. Abrams, S. Cruse, P., Kunze, J. (2008). Preservation is not a place. *International Journal of Digital Curation 1*(4). Retrieved August 1, 2013 from the WWW: http://www.ijdc.net/index.php/ijdc/article/view/98/73

   Description: An early snapshot of CDL

3. Acord, S., & Harley, D. (in press). Credit, time, and personality. *New Media and Society*. Retrieved August 1, 2013 from the WWW: http://nms-theme.ehumanities.nl/manuscript/credit-time-and-personality-acord-and-harley

   Abstract: We discuss the scholarly communication life cycle and examine the needs and values that drive academic behaviors, particularly within the early stages of sharing in-progress work. Second, we describe the significant tensions and obstacles to change in these practices as experienced by individual scholars across disciplines, specifically as they relate to receiving credit, managing finite time, and individual personality traits. By situating larger discussions about the future of scholarly communication in the everyday life of scholars, we argue that building continuity within disciplinary culture between conventional and new scholarly communication practices will be the key to the success of new initiatives.

4. Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. A. (2011). Public availability of published research data in high-impact journals. *PLoS ONE 6*(9), e24357. doi:10.1371/journal.pone.0024357

5. Altman, M. (2008). A fingerprint method for verification of scientific data. In T. Sobh (Ed.), *Advances in Computer and Information Sciences and Engineering* (pp. 311-316). Netherlands: Springer. Retrieved August 1, 2013 from the WWW: http://thedata.org/publications/fingerprint-method-verification-scientific-data

   Abstract: This article discusses an algorithm (called "UNF") for verifying digital data matrices. This algorithm is now used in a number of software packages and digital library projects. We discuss the details of the algorithm, and offer an extension for normalization of time and duration data.

6. Altman, M., & Crabtree, J. (2011). Using the SafeArchive system: TRAC-based auditing of LOCKSS. *Archiving 2011*, *7*, 165-170. Retrieved August 1, 2013 from the WWW: http://thedata.org/publications/using-safearchive-system-trac-based-auditing-lockss

   Abstract: The goals of SafeArchive are to make distributed replication easier, and to automate compliance with formal replication and storage policies. In this article, we describe the process of automated archival policy auditing in detail. First, we provide an overview of the SafeArchive system and we describe how a curator can use the tools to generate an archival policy schema and monitor it, simply. Second we identify specific TRAC criteria that can be verified automatically, and additional criteria that can be supported through integrated documentation. Third, we discuss the technical implementation of the system including the policy schema; how information used in the auditing process is obtained from a set of LOCKSS peers without modifying the LOCKSS trust model or configuration; and how the software is organized into components.

7. Altman, M., & King, G. (2007). A proposed standard for the scholarly citation of quantitative data. *D-Lib Magazine*, *13*(3/4). Retrieved August 1, 2013 from the WWW: http://gking.harvard.edu/files/abs/cite-abs.shtml

   Abstract: Citations to numerical data should include, at a minimum, six required components. The first three components are traditional, directly paralleling print documents. They include the author(s) of the data set, the date the data set was published or otherwise made public, and the data set title. The other three are: a unique global identifier, a universal numeric fingerprint, and a bridge service. They are also designed to take advantage of the digital form of quantitative data.

8. Altman, M., & Klass, G. M. (2005). Current research in voting, elections, and technology. *Social Science Computer Review, 23*(3), 269-273. doi: 10.1177/0894439305275849

Abstract: The articles in this special issue raise and refine questions about our understanding of the use of, state of the art in, and challenges associated with voting and election technology, broadly conceived. Although researchers have yet to achieve consensus on the broad impact of information technology on our understanding of the practice of politics, the broad outlines of a research agenda are emerging. In this overview, we discuss the current work and identify important research questions that remain to be addressed.

9. Altman, M., & Rogerson, K. (2008). Open research questions on information and technology in global and domestic politics – Beyond "E-." *PS: Political Science & Politics, 41*, 835-837. doi:http://dx.doi.org/10.1017/S1049096508081109

Abstract: Accelerating technological change is one of the defining characteristics of this era. And the intersection of information, technology, and politics is a constantly changing arena. Technological change can provide the subject for political debate, such as in the controversy over electronic voting (see Tokaji 2005); affect the means by which politics is conducted, such as in the use of information technologies to provide government services and collect regulatory feedback (see Fountain 2001; West 2005; and Mayer-Schonberger and Lazer 2007); or challenge our understanding of political theories and concepts, such as the meaning of privacy and of the public sphere (see Etzioni 2000 and Sunstein 2007 on the meaning of privacy and the compartmentalization of "public" speech, Bimber 2003 on the effect of information technologies on democracy, and Benkler 2006 on the reinterpretation of the public sphere). Each of these perspectives is visible locally, regionally, nationally, and globally.

10. Altman, M., Adams, M., Crabtree, J., Donakowski, D., Maynard, M., Pienta, A., & Young, C. (2009). Digital preservation through archival collaboration: The data preservation alliance for the social sciences. *The American Archivist, 72*(1), 170-184. Retrieved August 1, 2013 from the WWW: http://archivists.metapress.com/content/EU7252LHNRP7H188

Abstract: The Data Preservation Alliance for the Social Sciences (Data-PASS) is a partnership of five major U.S. institutions with a strong focus on archiving social science research. The Library of Congress supports the partnership through its National Digital Information Infrastructure and Preservation Program (NDIIPP). The goal of Data-PASS is to acquire and preserve data from opinion polls, voting records, large-scale surveys, and other social science studies at risk of being lost to the research community. This paper discusses the agreements, processes, and infrastructure that provide a foundation for the collaboration.

11. Altman, M., Andreev, L., Diggory, M., King, G., Sone, A., Verba, S., Kiskis, D. L., et al. (2001). A digital library for the dissemination and replication of quantitative social science research: The virtual data center. *Social Science Computer Review, 19*(4), 458-470. Retrieved August 1, 2013 from the WWW: http://www.box.net/shared/d3cf8u0gtyml2nqq3u2f

Abstract: The Virtual Data Center (VDC) software is a comprehensive, open-source, digital library system designated to help curators and researchers face the challenges of sharing and disseminating research data in an increasingly distributed world. The VDC is also a first step toward better citation of data. Current citations of data are typically ad hoc, fragile, and shallow. Ultimately, digital libraries such as the VDC will serve to make citations more robust and research more replicable.

12. Amos, H. (2011). Rsquared: researching the researchers. A study into how the researchers at the University of New South Wales use and share research data." *31st Annual IATUL Conference*. Retrieved August 1, 2013 from the WWW: http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1000&context=iatul2010

Abstract: This paper presents a research study of data usage, creation and sharing within different research communities at UNSW. The study identifies emerging data usage and management needs within the e-research life cycle of diverse research communities. Comparison is made with the outcomes of other studies that have examined e-researcher work practices in relation to their data. The paper examines the findings to understand what role researchers see libraries having, and discusses the development of a framework that libraries can use to support the curation and management of data and the development of tools and library support services that can be used across disciplines.

13. Anderegg, W., Prall, J., Harold, J., & Schneider, S. (2010). Expert credibility in climate change. *Proceedings of the National Academy of Science.* doi:10.1073/pnas.1003187107

Abstract: Here, we use an extensive dataset of 1,372 climate researchers and their publication and citation data to show that (i) 97–98% of the climate researchers most actively publishing in the field surveyed here support the tenets of ACC outlined by the Intergovernmental Panel on Climate Change, and (ii) the relative climate expertise and scientific prominence of the researchers unconvinced of ACC are substantially below that of the convinced researchers.

14. Anderson, R. G., Greene, W. H., McCullough, B. D., & Vinod, H. D. (2008). The role of data/code archives in the future of economic research. *Journal of Economic Methodology, 15*(1), 99-119. doi:10.1080/13501780801915574

15. Artz, D., & Gil, Y. (2007). A survey of trust in computer science and the semantic web. *Journal Web Semantics: Science, Services and Agents on the World Wide Web Archive. 5*(2), 58-71. doi:10.1016/j.websem.2007.03.002

Abstract: In computer science, trust is a widely used term whose definition differs among researchers and application areas. Trust is an essential component of the vision for the Semantic Web, where both new problems and new applications of trust are being studied. This paper gives an overview of existing trust research in computer science and the Semantic Web.

16. Autodesk Geospatial. (2007). Best practice for managing geospatial data. Retrieved August 1, 2013 from the WWW: http://www.gisperfect.com/res/AutocadMAP/best_practices.pdf

Description: Stage 1: AutoCAD or AutoCAD LT was used to create maps by engineers and drafting technicians, Stage 2: AutoCAD Map 3D used to create and edit geospatial data, Stage 3: AutoCAD Map 3D + FDO access multiple data sources, Stage 4: Spatial Databases extends the use of information-security and scalability, multiple users and sophisticated data models, Stage 5: Topobase and other applications are used in different departments in an enterprise. Managing spatial data using AutoCAD 3D

17. Ball, A., & Duke, M. (2011). Data citation and linking. *Digital Curation Centre*. Retrieved August 1, 2013 from the WWW: http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/data-citation-and-linking

18. Bechhofer, S., De Roure, D., Gamble, M. Goble, C., & Buchan, I. (2010). Research objects: Towards exchange and reuse of digital knowledge. In *The Future of the Web for Collaborative Science (FWCS 2010),* Raleigh, NC. Retrieved August 1, 2013 from the WWW: http://imageweb.zoo.ox.ac.uk/pub/2010/Proceedings/FWCS2010/05/Paper5.pdf

Abstract: What will researchers be publishing in the future? Whilst there is little question that the Web will be the publication platform, as scholars move away from paper towards digital content, there is a need for mechanisms that support the production of self-contained units of knowledge and facilitate the publication, sharing and reuse of such entities. In this paper we discuss the notion of research objects, semantically rich aggregations of resources that can possess some scientific intent or support some research objective. We present a number of principles that we expect such objects and their associated services to follow.

19. Bell, G., Hey, T., & Szalay, A. (2009). Beyond the data deluge. *Science, 323*(5919), 1297-1298. doi:10.1126/science.1170411

20. Bernstein, H. J., Folk, M. J., Benger, W., Dougherty, M. T., Eliceiri, K. W., & Schnetter, E. (2011). Communicating scientific data from the present to the future. Paper contributed to NSF Workshop, *Research Data Lifecycle Management*. Retrieved August 1, 2013 from the WWW: http://www.columbia.edu/~rb2568/rdlm/Bernstein_Dowling_RDLM2011.pdf

21. Bollen, J., Van de Sompel, H. (2006). An architecture for the aggregation and analysis of scholarly usage data. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 298-307). New York, NY: ACM. Retrieved August 1, 2013 from the WWW: http://arxiv.org/abs/cs/0605113

Abstract: Although recording of usage data is common in scholarly information services, its exploitation for the creation of value-added services remains limited due to concerns regarding, among others, user privacy, data validity, and the lack of accepted standards for the representation, sharing and aggregation of

usage data. This paper presents a technical, standards-based architecture for sharing usage information, which we have designed and implemented. In this architecture, OpenURL-compliant linking servers aggregate usage information of a specific user community as it navigates the distributed information environment that it has access to. This usage information is made OAI-PMH harvestable so that usage information exposed by many linking servers can be aggregated to facilitate the creation of value-added services with a reach beyond that of a single community or a single information service. This paper also discusses issues that were encountered when implementing the proposed approach, and it presents preliminary results obtained from analyzing a usage data set containing about 3,500,000 requests aggregated by a federation of linking servers at the California State University system over a 20 month period.

22. Bollen, J., & Van de Sompel, H. (2006). Mapping the structure of science through usage. *Scientometrics, 69*(2), 227-258. Retrieved August 1, 2013 from the WWW: http://public.lanl.gov/herbertv/papers/Papers/2006/SCIENTObollen_map.pdf

    Abstract: Science has traditionally been mapped on the basis of authorship and citation data. Due to publication and citation delays such data represents the structure of science as it existed in the past. We propose to map science by proxy of journal relationships derived from usage data to determine research trends as they presently occur. This mapping is performed by applying a principal components analysis superimposed with a k-means cluster analysis on networks of journal relationships derived from a large set of article usage data collected for the Los Alamos National Laboratory research community. Results indicate that meaningful maps of the interests of a local scientific community can be derived from usage data. Subject groupings in the mappings corresponds to Thomson ISI subject categories. A comparison to maps resulting from the analysis of 2003 Thomson ISI Journal Citation Report data reveals interesting differences between the features of local usage and global citation data.

23. Bollen, J., & Van de Sompel, H. (2008). Usage impact factor: The effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology, 59*(1), 136-149. Retrieved August 1, 2013 from the WWW: http://arxiv.org/pdf/cs.DL/0610154.pdf

    Abstract: There exist ample demonstrations that indicators of scholarly impact analogous to the citation-based ISI Impact Factor can be derived from usage data. However, contrary to the ISI IF which is based on citation data generated by the global community of scholarly authors, so far usage can only be practically recorded at a local level leading to community-specific assessments of scholarly impact that are difficult to generalize to the global scholarly community. We define a journal Usage Impact Factor which mimics the definition of the Thomson Scientific's ISI Impact Factor. Usage Impact Factor rankings are calculated on the basis of a large-scale usage data set recorded for the California State University system from 2003 to 2005. The resulting journal rankings are then compared to Thomson Scientific's ISI Impact Factor which is used as a baseline indicator of general impact. Our results indicate that impact as derived from California State University usage reflects the particular scientific and demographic characteristics of its communities.

24. Bollen, J., Van de Sompel, H., Hagberg, A. Bettencourt, L., Chute, R., Rodriguez, M., & Balakireva, L. (2009). Clickstream data yields high-resolution maps of science. *PLoS One, 4*. doi:10.1371/journal.pone.0004803

    Abstract: Intricate maps of science have been created from citation data to visualize the structure of scientific activity. However, most scientific publications are now accessed online. Scholarly web portals record detailed log data at a scale that exceeds the number of all existing citations combined. Such log data is recorded immediately upon publication and keeps track of the sequences of user requests (clickstreams) that are issued by a variety of users across many different domains. Given these advantages of log datasets over citation data, we investigate whether they can produce high-resolution, more current maps of science.

25. Bollen, J., Van de Sompel, H., HagBerg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *Cornell University Library*. doi:10.1371/journal.pone.0006022

    Abstract: The impact of scientific publications has traditionally been expressed in terms of citation counts. However, scientific activity has moved online over the past decade. To better capture scientific impact in the digital era, a variety of new impact measures has been proposed on the basis of social network analysis and usage log data. Here we investigate how these new measures relate to each other, and how accurately

and completely they express scientific impact. We performed a principal component analysis of the rankings produced by 39 existing and proposed measures of scholarly impact that were calculated on the basis of both citation and usage log data. Our results indicate that the notion of scientific impact is a multi-dimensional construct that can not be adequately measured by any single indicator, although some measures are more suitable than others. The commonly used citation Impact Factor is not positioned at the core of this construct, but at its periphery, and should thus be used with caution.

26. Bollen, J., Rodriguez, M., & Van de Sompel, H. (2006). Journal status. *Scientometrics, 69*(3), 669-687. doi:10.1007/s11192-006-0176-z

    Abstract: The status of an actor in a social context is commonly defined in terms of two factors: the total number of endorsements the actor receives from other actors and the prestige of the endorsing actors. These two factors indicate the distinction between popularity and expert appreciation of the actor, respectively. We refer to the former as popularity and to the latter as prestige. These notions of popularity and prestige also apply to the domain of scholarly assessment. The ISI Impact Factor (ISI IF) is defined as the mean number of citations a journal receives over a 2 year period. By merely counting the amount of citations and disregarding the prestige of the citing journals, the ISI IF is a metric of popularity, not of prestige. We demonstrate how a weighted version of the popular PageRank algorithm can be used to obtain a metric that reflects prestige. We contrast the rankings of journals according to their ISI IF and their weighted PageRank, and we provide an analysis that reveals both significant overlaps and differences. Furthermore, we introduce the Y-factor which is a simple combination of both the ISI IF and the weighted PageRank, and find that the resulting journal rankings correspond well to a general understanding of journal status.

27. Bollen, J., Van de Sompel, H., Smith, J., & Luce, R. (2005). Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing & Management, 41*(6), 1419-1440. Retrieved August 1, 2013 from the WWW: http://arxiv.org/abs/cs.DL/0503007

    Abstract: We generated networks of journal relationships from citation and download data, and determined journal impact rankings from these networks using a set of social network centrality metrics. The resulting journal impact rankings were compared to the ISI IF. Results indicate that, although social network metrics and ISI IF rankings deviate moderately for citation-based journal networks, they differ considerably for journal networks derived from download data. We believe the results represent a unique aspect of general journal impact that is not captured by the ISI IF. These results furthermore raise questions regarding the validity of the ISI IF as the sole assessment of journal impact, and suggest the possibility of devising impact metrics based on usage information in general.

28. Bollen, J., Van de Sompel, H., & Rodriguez, M. (2008). Towards usage-based impact metrics: First results from the MESUR project." In *Proceedings of the Joint Conference on Digital Libraries 2008,* Pittsburgh, PA. Retrieved August 1, 2013 from the WWW: http://arxiv.org/abs/0804.3791

    Abstract: Scholarly usage data holds the potential to be used as a tool to study the dynamics of scholarship in real time, and to form the basis for the definition of novel metrics of scholarly impact. However, the formal groundwork to reliably and validly exploit usage data is lacking, and the exact nature, meaning and applicability of usage-based metrics is poorly understood. The MESUR project funded by the Andrew W. Mellon Foundation constitutes a systematic effort to define, validate and cross-validate a range of usage-based metrics of scholarly impact. MESUR has collected nearly 1 billion usage events as well as all associated bibliographic and citation data from significant publishers, aggregators and institutional consortia to construct a large-scale usage data reference set. This paper describes some major challenges related to aggregating and processing usage data, and discusses preliminary results obtained from analyzing the MESUR reference data set. The results confirm the intrinsic value of scholarly usage data, and support the feasibility of reliable and valid usage-based metrics of scholarly impact.

29. Borgman, C. (2009). The future is now: A call to action for the humanities. *DHQ 3*(4). Retrieved August 1, 2013 from the WWW: http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html

30. Borgman, C. (2011). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* (pp. 1-40). doi:http://dx.doi.org/10.2139/ssrn.1869155

    Abstract: This article explores the complexities of data, research practices, innovation, incentives, economics, intellectual property, and public policy associated with the data sharing conundrum – "an

intricate and difficult problem." Research data take many forms, are collected for many purposes, via many approaches, and often are difficult to interpret once removed from their initial context. Rationales for sharing data vary along two dimensions: whether motivated by research concerns or by leveraging public investments, and whether intended to serve the interests of researchers who produce data or the interests of potential re-users of data. Four rationales for sharing research data are identified and positioned on these dimensions. Researchers' incentives to share their data depend not only on these rationales, but on characteristics of their data and research practices, funding agency policies, and resources for data management. Much more is understood about why researchers do not share data than about when, why, and how researchers do share data, or about when, how, and why researchers or the public reuse data. The model and research agenda are illustrated with examples from the sciences, social sciences, and humanities.

31. Bose, R., & Frew, J. (2005). Lineage retrieval for scientific data processing. *ACM Computing Surveys, 37*(1), 1-28. doi:10.1145/1057977.1057978

    Abstract: Scientific research relies as much on the dissemination and exchange of data sets as on the publication of conclusions. Accurately tracking the lineage (origin and subsequent processing history) of scientific data sets is thus imperative for the complete documentation of scientific work. Researchers are effectively prevented from determining, preserving, or providing the lineage of the computational data products they use and create, however, because of the lack of a definitive model for lineage retrieval and a poor fit between current data management tools and scientific software. Based on a comprehensive survey of lineage research and previous prototypes, we present a metamodel to help identify and assess the basic components of systems that provide lineage retrieval for scientific data products.

32. Bourne, P. (2005). Will a biological database be different from a biological journal? *PLoS Computational Biology, 1*, e34. doi:10.1371/journal.pcbi.0010034

33. Brase, J. (2004). Using digital library techniques: Registration of scientific primary data. *Research and Advanced Technology for Digital Libraries Lecture Notes in Computer Science, 3232*, 484-494. doi:10.1007/978-3-540-30230-8_44

    Abstract: Registration of scientific primary data, to make these data citable as a unique piece of work and not only a part of a publication, has always been an important issue. With the new digital library techniques, it is finally made possible. In the context of the project Publication and Citation of Scientific Primary Data founded by the German research foundation (DFG) the German national library of science and technology (TIB) has become the first registration agency worldwide for scientific primary data. The datasets receive unique DOIs and URNs as citable identifiers and all relevant metadata information is stored at the online library catalogue. Registration has started for the field of earth science, but will be widened for other subjects in 2005. In this paper we will give you a quick overview about the project and the registration of primary data.

34. Brase, J., Farquhar, A., Gastl, A., Gruttemeier, H., Heijne, M., Heller, A., Hitson, B., Johnson, L., et al. (2009). Numeric data: Citation techniques and integration with text. *General format.* Retrieved August 1, 2013 from the WWW: http://www.icsti.org/IMG/pdf/Numeric_Data_FINAL_report.pdf

    Abstract: The scientific and information communities have largely mastered the presentation of, and linkages between, text-based electronic information by assigning persistent identifiers to give scientific literature unique identities and accessibility. Knowledge, as published through scientific literature, is often the last step in a process originating from scientific research data. Today scientists are using simulation, observational, and experimentation techniques that yield massive quantities of research data. These data are analysed, synthesised, interpreted, and the outcome of this process is generally published as a scientific article. Access to the original data as the foundation of knowledge has become an important issue throughout the world and different projects have started to find solutions. Global collaboration and scientific advances could be accelerated through broader access to scientific research data. In other words, data access could be revolutionized through the same technologies used to make textual literature accessible. The most obvious opportunity to broaden visibility of and access to research data is to integrate its access into the medium where it is most often cited: electronic textual information. Besides this opportunity, it is important, irrespective of where they are cited, for research data to have an internet identity.

35. Brase, J. Farquhar, A., Gastl, A., Gruttemeier, H., Heijne, M., Heller, A., Piguet, A., et al. (2009). Approach for a joint global registration for research data. *Information Services & Use*, *29*, 13–27. doi: 10.3233/ISU-2009-0595

Abstract: Data access could be revolutionized through the same technologies used to make textual literature accessible. The most obvious opportunity to broaden visibility of and access to research data is to integrate its access into the medium where it is most often cited: electronic textual information. Besides this opportunity, it is important, irrespective of where they are cited, for research data to have an internet identity. Since 2005, the German National Library of Science and Technology (TIB) has offered a successful Digital Object Identifier (DOI) registration service for persistent identification of research data. In this white paper we discuss the possibilities to open this registration to a global consortium of information institutes and libraries.

36. Buneman, P. (2006). How to cite curated databases and how to make them citable. *Proceedings of the 18th International Conference on Scientific and Statistical Database Management*, Vienna, July 2006. Retrieved August 1, 2013 from the WWW: http://homepages.inf.ed.ac.uk/opb/papers/ssdbm2006.pdf

37. Buneman, P., & Silvello, G (2010). A rule-based citation system for structured and evolving datasets. *IEEE Data Engineering Bulletin, 33*(3), 33-41. Retrieved August 1, 2013 from the WWW: http://sites.computer.org/debull/A10sept/buneman.pdf

Abstract: We consider the requirements that a citation system must fulfill in order to cite structured and evolving data sets. Such a system must take into account variable granularity, context and the temporal dimension. We look at two examples and discuss the possible forms of citation to these data sets. We also describe a rule-based system that generates citations which fulfill these requirements.

38. Buneman, P., Khanna, S. & Wang-Chiew, T. (2001).Why and where: A characterization of data provenance. In *Database Theory – Proceedings of the ICDT 2001*, 316-330. Berlin: Springer-Verlag. Retrieved August 1, 2013 from the WWW: http://db.cis.upenn.edu/DL/whywhere.pdf

39. Butler, D. (2012). Scientists: Your number is up: ORCID scheme will give researchers unique identifiers to improve tracking of publications. *Nature, 485*, 564. doi:10.1038/485564a

40. Callaghan, C., Donegan, S, Pepler, S. Thorley, M., Cunningham, N., Kirsch, P., Ault, L., et al. (2012). Making data a first class scientific output: Data citation and publication by NERC's environmental data centres. *International Journal of Digital Curation, 7*(1), 107-113. doi:10.2218/ijdc.v7i1.218

Abstract: The NERC Science Information Strategy Data Citation and Publication project aims to develop and formalise a method for formally citing and publishing the datasets stored in its environmental data centres. It is believed that this will act as an incentive for scientists, who often invest a great deal of effort in creating datasets, to submit their data to a suitable data repository where it can properly be archived and curated. Data citation and publication will also provide a mechanism for data producers to receive credit for their work, thereby encouraging them to share their data more freely.

41. Campbell, E.G., & Bendavid, E. (2003). Data-sharing and data-withholding in genetics and the life sciences: Results of a national survey of technology transfer officers. *Journal of Health Care Law and Policy, 6*(2), 241. Retrieved August 1, 2013 from the WWW: http://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=1216&context=jhclp

Abstract: The completion of a working draft of the human genome sequence two years ago will, no doubt, prove to be an integral chapter in a story of extraordinary technological achievement- a story based on the continued revelation of genetic information. The public debate aside, the federal courts, principally the U.S. Court of Appeals for the Federal Circuit, and the U.S. Patent and Trademark Office have both attempted to provide guidance on the intellectual property rights that might impact such matters involving the human genome and other genetic data. These efforts, however, have met with lackluster support at best from patent law practitioners and other commentators, as well as the general public. Moreover, attempts to obtain patent protection for early stage research products may negatively impact scientific progress. Given the rapidity with which technology will be available to affect whole genomic sequencing over the next decade, new models must also emerge to engage these capabilities within the health care regime, and to guard against exploitation by those "with access" to the detriment of the individual.

42. Case, D. O., & Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, 51, 635–645. doi:10.1002/(SICI)1097-4571(2000)51:7<635::AID-ASI6>3.0.CO;2-H

43. Chavan, V., & Ingwersen, P. (2009). Towards a data publishing framework for primary biodiversity data: Challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics, 10* (Suppl. 14), S2. doi:10.1186/1471-2105-10-S14-S2

    Abstract: Currently primary scientific data, especially that dealing with biodiversity, is neither easily discoverable nor accessible. Amongst several impediments, one is a lack of professional recognition of scientific data publishing efforts. A possible solution is establishment of a '*Data Publishing Framework'* which would encourage and recognise investments and efforts by institutions and individuals towards management, and publishing of primary scientific data potentially on a par with recognitions received for scholarly publications.

44. Cheney, J., Chiticariu, L., & Tan,W.-T. (2009). Provenance in databases: Why, how, and where. *Foundations and Trends® in Databases, 1*(4), 379-474. doi: 10.1561/1900000006

    Abstract: Different notions of provenance for database queries have been proposed and studied in the past few years. In this article, we detail three main notions of database provenance, some of their applications, and compare and contrast amongst them. Specifically, we review why, how, and where provenance, describe the relationships among these notions of provenance, and describe some of their applications in confidence computation, view maintenance and update, debugging, and annotation propagation.

45. CIESIN Columbia University (2005). Data model for managing and preserving geospatial electronic records. *General format.* Retrieved August 1, 2013 from the WWW: http://www.ciesin.columbia.edu/ger/DataModelV1_20050620.pdf

    Description: The article consists of a data model for managing and preserving Geospatial records and how to improve capabilities of systems already implemented. It has description of the model, UML diagram, data dictionary and capability to crosswalk with other schemas.

46. Cole, F. (2008). Taking "data" (as a topic): The working policies of indifference, purification and differentiation. *Association for Information Systems Electronic Library.* Retrieved August 1, 2013 from the WWW: http://aisel.aisnet.org/acis2008/79/

    Abstract: The recent surge of interest in e-science presents an opportune moment to re-examine the fundamental idea of "data". This paper explores this topic by reporting on the different ways in which the idea of data is handled across many disciplines. From the accounts various disciplines themselves provide, these ways can be portrayed as the pursuit of three broad policies. The first policy is one of Indifference, which assumes the coherence of the data-concept, so that there is no need to explicate it further. The second policy is Purification, which identifies the essential characteristics of data according to the conventions of a particular discipline, with other modes systematically suppressed. The third policy allows for the Differentiation that is evident in the manifestations of data in various disciplines that utilise information systems. Greater appreciation among information professionals of the alternative approaches to data hopefully will enhance policy formulation and systems design.

47. Cook, R., Olson, R., Kancriruk, P., & Hook, L. (2000). Best practices for preparing ecological and ground-based data sets to share and archive. *Environmental Sciences Division, Oak Ridge National Laboratory*. Retrieved August 1, 2013 from the WWW: www.daac.ornl.gov/DAAC/PI/bestprac.html#prac2

    Description: Provides guidelines to improve usability and allow sharing of datasets with other researchers. The seven best practices are: Assign Descriptive File Names, Use Consistent and Stable File Formats for Tabular and Image Data, Define the Contents of Your Data Files, Use Consistent Data Organization, Perform Basic Quality Assurance, Assign Descriptive Data Set Titles, Provide Documentation

48. Costello, M. J. (2009). Motivating online publication of data. *Bioscience, 59*(5), 418-427. doi:10.1525/bio.2009.59.5.9

    Abstract: Despite policies and calls for scientists to make data available, this is not happening for most environmental- and biodiversity-related data because scientists' concerns about these efforts have not been answered and initiatives to motivate scientists to comply have been inadequate. Many of the issues

regarding data availability can be addressed if the principles of "publication" rather than "sharing" are applied. However, online data publication systems also need to develop mechanisms for data citation and indices of data access comparable to those for citation systems in print journals.

49. Cragin, M. H., Palmer, C. L., Carlson, J.R., &Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society, 368*(1926), 4023-4038. doi:10.1098/rsta.2010.0165

    Abstract: Results are presented from the Data Curation Profiles project research, on who is willing to share what data with whom and when. Emerging from scientists' discussions on sharing are several dimensions suggestive of the variation in both what it means 'to share' and how these processes are carried out. This research indicates that data curation services will need to accommodate a wide range of subdisciplinary data characteristics and sharing practices. As part of a larger set of strategies emerging across academic institutions, institutional repositories (IRs) will contribute to the stewardshipmobilization of scientific research data for e-Research and learning. There will be particular types of data that can be managed well in an IR context when characteristics and practices are well understood. Findings from this study elucidate scientists' views on 'sharable' forms of data—the particular representation that they view as most valued for reuse by others within their own research areas—and the anticipated duration for such reuse. Reported sharing incidents that provide insights into barriers to sharing and related concerns on data misuse are included.

50. Crosas, M. (2011). The dataverse network®: An open-source application for sharing, discovering and preserving data. *D-Lib Magazine, 17*(½). doi:10.1045/january2011-crosas

    Abstract: The Dataverse Network is an open-source application for publishing, referencing, extracting and analyzing research data. The main goal of the Dataverse Network is to solve the problems of data sharing through building technologies that enable institutions to reduce the burden for researchers and data publishers, and incentivize them to share their data. By installing Dataverse Network software, an institution is able to host multiple individual virtual archives, called "dataverses" for scholars, research groups, or journals, providing a data publication framework that supports author recognition, persistent citation, data discovery and preservation. Dataverses require no hardware or software costs, nor maintenance or backups by the data owner, but still enable all web visibility and credit to devolve to the data owner.

51. DataCite. (2011). DataCite metadata scheme for the publication and citation of research data, Version 2.2. doi:10.5438/0005

52. Davidson, S. B., & Freire, J. (2008). Provenance and scientific workflows: challenges and opportunities. *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. Retrieved July 30, 2013 from the WWW: http://bigdata.poly.edu/~juliana/pub/freire-tutorial-sigmod2008.pdf

53. Dinkelmann, K., Edwards, M., Fry, J., Humphrey, C., Nakao, R., & Thomas, W. (2009). Work flows - data discovery and dissemination: User perspective." *Data Documentation Initiative, Working Paper Series.* doi: http://dx.doi.org/10.3886/DDIBestPractices02

    Description: Describes the best practices for metadata producers to provide end users with the resources for data discovery and dissemination. Citation not addressed

54. Downs, R. R., & Chen, R.S. (2005). Organizational needs for managing and preserving geospatial data and related electronic records. *Data Science Journal, 4*, 255-271. Retrieved August 1, 2013 from the WWW: https://www.jstage.jst.go.jp/article/dsj/4/0/4_0_255/_pdf

    Abstract: Government agencies and other organizations are required to manage and preserve records that they create and use to facilitate future access and reuse. The increasing use of geospatial data and related electronic records presents new challenges for these organizations, which have relied on traditional practices for managing and preserving records in printed form. This article reports on an investigation of current and future needs for managing and preserving geospatial electronic records on the part of local- and state-level organizations in the New York City metropolitan region. It introduces the study and describes organizational needs observed, including needs for organizational coordination and inter-organizational cooperation throughout the entire data lifecycle.

55. Duerr, R, Downs, R., Tilmes, C., Barkstrom, B., Lenhardt, W., Glassy, J., Bermudez, L., & Slaughter, P. (2011). On the utility of identification schemes for digital earth science data: An assessment and recommendations." *Earth Science Informatics, 4*(3),139-160. doi:http://dx.doi.org/10.1007/s12145-011-0083-6

Abstract: In recent years, a number of data identification technologies have been developed which purport to permanently identify digital objects. In this paper, nine technologies and systems for assigning persistent identifiers are assessed for their applicability to Earth science data (ARKs, DOIs, XRIs, Handles, LSIDs, OIDs, PURLs, URIs/URNs/URLs, and UUIDs). The evaluation used four use cases that focused on the suitability of each scheme to provide Unique Identifiers for Earth science data objects, to provide Unique Locators for the objects, to serve as Citable Locators, and to uniquely identify the scientific contents of data objects if the data were reformatted. Of all the identifier schemes assessed, the one that most closely meets all of the requirements for an Unique Identifier is the UUID scheme. Any of the URL/URI/IRI-based identifier schemes assessed could be used for Unique Locators. Since there are currently no strong market leaders to help make the choice among them, the decision must be based on secondary criteria. While most publications now allow the use of URLs in citations, so that all of the URL/URI/IRI based identification schemes discussed in this paper could potentially be used as a Citable Locator, DOIs are the identification scheme currently adopted by most commercial publishers. None of the identifier schemes assessed here even minimally address identification of scientifically identical numerical data sets under reformatting.

56. Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences in the United States of America, 109*(42), 17028-17033. doi:10.1073/pnas.1212247109

57. Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17*(1), 120-128. doi:10.1037/a0024445

58. Fitzgerald, A., & Pappalardo, K. (2007). Building the infrastructure for data access and reuse in collaborative research: An Analysis of the Legal Context. *The OAK Law Project, Canberra, Australia.* Retrieved August 1, 2013 from the WWW: http://eprints.qut.edu.au/8865/1/8865.pdf

Description: This Report examines the broad legal framework within which research data is generated, managed, disseminated and used. The background to the Report is the growing support for systems that enable research data generated in publicly-funded research projects to be made available for access and use by others in the research community.

59. Freire, J., Koop, D., Santos, E., & Silva, C. (2008). Provenance for computational tasks: A survey. *Computing Science and Engineering, 10*(3), 11-21. Retrieved August 1, 2013 from the WWW: http://vgc.poly.edu/~juliana/pub/freire-cise2008.pdf

Abstract: The problem of systematically capturing and managing provenance for computational tasks has recently received significant attention because of its relevance to a wide range of domains and applications. The authors give an overview of important concepts related to provenance management, so that potential users can make informed decisions when selecting or designing a provenance solution.

60. Friends of the Chair Group on Integrated Economic Statistics (2007 June). Session 3(c): Dissemination standards (data and metadata), data exchange and revision policy. Retrieved August 1, 2013 from the WWW: http://www.bfs.admin.ch/bfs/portal/en/index/institutionen/statistikaemter_in/03/02.parsys.0021.downloadList.00211.DownloadFile.tmp/disseminationstandardsdataandmetadatadataexchangeandrevisionpolicyoecd3c.pdf

61. Fry, J., Houghton, J., Lockyer, S., Oppenheim, C., & Rasmussen, B., (2008). Identifying benefits arising from the curation and open sharing of research data produced by UK higher education and research institutes. Retrieved August 1, 2013 from the WWW: http://ie-repository.jisc.ac.uk/279/

Description: A review study was commissioned from UKOLN on how data is managed in the UK. The aim of the project is to identify the benefits of the curation and open sharing of research data, using quantitative and qualitative methods. Citation is not mentioned.

62. Gantz, J., & Reinsel, D. (2010). The digital universe decade – Are you ready? Retrieved August 1, 2013 from the WWW: http://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf.

Media here: http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm

63. Gantz, J., Chute, C., Manfrediz, A., Minton, S., Reinsel, D., Schlichting, W., & Toncheva , A. (2008). The diverse and exploding digital universe. An updated forecast of worldwide information growth through 2011. Retrieved August 1, 2013 from the WWW: http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf

Description Mainly focusing on data growth.

64. Gibbs, H. (2007). DISC-UK DataShare: State-of-the-art review. *Data Share Project*. Retrieved August 1, 2013 from the WWW: http://www.disc-uk.org/docs/state-of-the-art-review.pdf

65. Gold, A. (2007). Cyberinfrastructure, data, and libraries, part 1. *D-Lib Magazine*, *13*(9/10). Retrieved July 30, 2013 from the WWW: http://www.dlib.org/dlib/september07/gold/09gold-pt1.html

66. Gonzales, E., Zhang X., Akahoshi Y., Murayama Y., Zettsu K. (2012, October). Data Citation Wiki for Harnessing Collective Intelligence on Document-to-Data Associations to Interdisciplinary Data Access. Paper presented at the 23rd International CODATA Conference, Taipei, China.

67. Green, A., Macdonald, S., & Rice, R. (2009). Policy-making for research data in repositories: A guide. *JISC funded DISC-UK Share Project*. doi:10.1787/603233448430

Description: This article discusses citation briefly in the context of access and reuse of data. No survey done.

68. Green, T. (2009). We need publishing standards for datasets and data tables. *OECD Publishing White Paper*, Retrieved from http://dx.doi.org/10.1787/603233448430

Description: This article advocates a slightly more verbose citation standard than Altman & King. (includes a comparison table for the two standards). In the new system being built by OECD, All the DOIs for the datasets and tables will be deposited with CrossRef, ready for other publishers to use.

69. Greenberg, S. (2009). How citation distortions create unfounded authority: analysis of a citation network *BMJ* 339:b2680. doi:10.1136/bmj.b2680

70. Griffiths, A. (2009). The publication of research data: Researcher attitudes and behaviors. *The International Journal of Digital Curation*, *4*(1), 46-56. Retrieved July 30, 2013 from the WWW: http://www.ijdc.net/index.php/ijdc/article/view/101/76

71. Gutmann, M., Abrahamson, M., Adams, M., Altman, M., Arms, C., Bollen, K., . . ., Young, C. (2009). From preserving the past to preserving the future: The data-PASS project and the challenges of preserving digital social science data. *Library Trends*, *57*(3), 315-337. Retrieved July 30, 2013 from the WWW: http://gking.harvard.edu/files/gking/files/GutAbrAda09.pdf

72. Haak, L. L.,Fenner, M.,Paglione, L.,Pentz, E., &Ratner, H. (2012). ORCID: A system to uniquely identify researchers. *Learned Publishing*, *25*, 259-264. doi:http://dx.doi.org/10.1087/20120404

73. Hakala, J. (2010). Persistent identifiers: An overview. *The KIM Technology Watch Report.* Retrieved August 1, 2013 from the WWW: http://metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/

Description: This article describes five persistent identifier systems (ARK, DOI, PURL, URN and XRI) and compares their functionality against the cool URIs. The aim is to provide an overview, not to give any kind of ranking of these systems.

74. Hamilton, E. (2007). The impact of survey data: Measuring success. *Journal of the American Society for Information Science and Technology*, *58*(2), 190–199. doi:10.1002/asi.20458

Abstract: Large national social surveys are expensive to conduct and to process into usable data files. The purpose of this article is to assess the impact of these national data sets on research using bibliometric measures. Peer-reviewed articles from research using numeric data files and documentation from the Canadian National Population Health Survey (NPHS) were searched in ISI's Web of Science and in Scopus

for articles citing the original research. This article shows that articles using NPHS data files and products have been used by a diverse and global network of scholars, practitioners, methodologists, and policy makers.

75. Harley, D., & Acord, S. (2011). *Peer review in academic promotion and publishing: Its meaning, locus, and future*. University of California, Berkeley: Center for Studies in Higher Education. Retrieved August 1, 2013 from the WWW: http://escholarship.org/uc/item/1xv148c8

    Abstract: The current phase of the project focuses on peer review in the Academy; this deeper look at peer review is a natural extension of our findings in Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines (Harley et al. 2010), which stressed the need for a more nuanced academic reward system that is less dependent on citation metrics, the slavish adherence to marquee journals and university presses, and the growing tendency of institutions to outsource assessment of scholarship to such proxies as default promotion criteria. This investigation is made urgent by a host of new challenges facing institutional peer review, such as assessing interdisciplinary scholarship, hybrid disciplines, the development of new online forms of edition making and collaborative curation for community resource use, heavily computational subdisciplines, large-scale collaborations around grand challenge questions, an increase in multiple authorship, a growing flood of low-quality publications, and the call by governments, funding bodies, universities, and individuals for the open access publication of taxpayer-subsidized research, including original data sets. This report includes (1) an overview of the state of peer review in the Academy at large, (2) a set of recommendations for moving forward, (3) a proposed research agenda to examine in depth the effects of academic status-seeking on the entire academic enterprise, (4) proceedings from the workshop on the four topics noted above, and (5) four substantial and broadly conceived background papers on the workshop topics, with associated literature reviews.

76. Heery, R. (2009). *Digital repositories roadmap review: Towards a vision for research and learning in 2013.* Retrieved August 1, 2013 from the WWW: http://www.jisc.ac.uk/media/documents/themes/infoenvironment/reproadmapreviewfinal.doc

    Description: Addresses citation metrics

77. Helliwell, J., & McMahon, B. (2010). The record of experimental science: Archiving data with literature. *Information services and use – Selected papers from the ICSTI interactive publications conference 2010, 30* (1/2), 31-37. doi: 10.3233/ISU-2010-0609

    Abstract: Crystallography is presented as a case study of a scientific discipline where the experimental data that underpin research results can be integrated into the scientific record. Among other advantages, this maximizes the degree of trust in science, since published results can thereby always be validated independently.

78. Helly, J., Elvins, T., Sutton, D. Martinez, S. Miller, S. Pickett, & Ellison, A. M. (2002). Controlled publication of digital scientific data. *CACM 45*(5), 97-101. doi:10.1145/506218.506222

    Description: This article discusses how to balance free and open access to scientific data with privileged access to new results by authors while protecting them from being scooped by competing interpretations of their own data.

79. Hey, T., & Trefethen, A. (2003). The data deluge: An e-science perspective. In Berman, F., Fox, G. C., & Hey, A. J. G. (Eds.), *Grid Computing - Making the Global Infrastructure a Reality* (pp. 809-824). New York, NY: Wiley. Retrieved August 1, 2013 from the WWW: http://eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf

    Abstract: This paper previews the imminent flood of scientific data expected from the next generation of experiments, simulations, sensors and satellites. In order to be exploited by search engines and data mining software tools, such experimental data needs to be annotated with relevant metadata giving information as to provenance, content, conditions and so on. The need to automate the process of going from raw data to information to knowledge is briefly discussed. The paper argues the case for creating new types of digital libraries for scientific data with the same sort of management services as conventional digital libraries in addition to other data-specific services. Some likely implications of both the Open Archives Initiative and e-Science data for the future role for university libraries are briefly mentioned. A substantial subset of this

e-Science data needs to archived and curated for long-term preservation. Some of the issues involved in the digital preservation of both scientific data and of the programs needed to interpret the data are reviewed. Finally, the implications of this wealth of e-Science data for the Grid middleware infrastructure are highlighted.

80. Hook, L., Vannan, A., Beaty, T., Cook, R., & Wilson, B. (2010). Best practices for preparing environmental data sets to share and archive 1. *Environmental Sciences Division*. Retrieved August 1, 2013 from the WWW: http://daac.ornl.gov/PI/BestPractices-2010.pdf

    Description: The most important practices that researchers could implement is to make their data sets ready to share with other researchers. These practices could be performed at any time during the preparation of the data set, but we suggest that researchers consider them before measurements are taken. The order of the practices is not necessarily sequential, as a researcher could provide draft data set metadata before any measurements are taken.

81. Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D., Kania, R., Schaeffer, M., St Pierre, S., Twigger, S., White, O., & Rhee, S. (2008). Big data: The future of biocuration. *Nature, 455*, 47-50. doi:10.1038/455047a

    Description: With the growth in the amount of biological data means that revolutionary measures are needed for data management, analysis and accessibility. Biocuration, the activity of organizing, representing and making biological information accessible to both humans and computers, has become an essential part of biological discovery and biomedical research.

82. Humphreys, M., Sanchez de la Sierra, R., & van der Windt, P. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, *21*(1), 1-20. doi:10.1093/pan/mps021

83. Jones, S. (2012). Developments in research funder data policy. *International Journal of Digital Curation*, 7(1), 114-125. doi:10.2218/ijdc.v7i1.219

    Abstract: This paper reviews developments in funders' data management and sharing policies, and explores the extent to which they have affected practice. The Digital Curation Centre has been monitoring UK research funders' data policies since 2008.There have been significant developments in subsequent years, most notably the joint Research Councils UK's Common Principles on Data Policy and the Engineering and Physical Sciences Research Council's Policy Framework on Research Data. This paper charts these changes and highlights shifting emphasizes in the policies. Institutional data policies and infrastructure are increasingly being developed as a result of these changes. While action is clearly being taken, questions remain about whether the changes are affecting practice on the ground.

84. Kethers, S., Shen X., Treloar, A.E., & Wilkinson, R. G. (2010). Discovering Australia's research data. In *JCDL '10: Proceedings of the 10th Annual Joint Conference on Digital Libraries* (pp. 345-348). New York, NY: ACM. Retrieved August 1, 2013 from the WWW: http://andrew.treloar.net/research/publications/jcdl2010/jcdl158-kethers.pdf

    Abstract: This paper argues that it is important to make it easier to find and access data that might be found in an institution, in a disciplinary data store, in a government department, or held privately. We explore how to meet ad hoc needs that cannot easily be supported by a disciplinary ontology, and argue that web pages that describe data collections with rich links and rich text are valuable. We describe the approach followed by the Australian National Data Service (ANDS) in making such pages available.

85. King, G. (2007). An introduction to the Dataverse network as an infrastructure for data sharing. *Sociological Methods & Research*, 36(2), 173-199. Retrieved August 1, 2013 from the WWW: http://gking.harvard.edu/gking/files/dvn.pdf

    Abstract: We introduce a set of integrated developments in web application software, networking, data citation standards, and statistical methods designed to put some of the universe of data and data sharing practices on somewhat firmer ground. We have focused on social science data, but aspects of what we have developed may apply more widely. The idea is to facilitate the public distribution of persistent, authorized, and verifiable data, with powerful but easy-to-use technology, even when the data are confidential or proprietary. We intend to solve some of the sociological problems of data sharing via technological means,

with the result intended to benefit both the scientific community and the sometimes apparently contradictory goals of individual researchers.

86. Kirlew, P. (2011). Life science data repositories in the publications of scientists and librarians. *Issues in Science and Technology Librarianship, 65*. doi:10.5062/F4X63JT2

87. Kowalczyk, S., & Shankar, K. Data sharing in the sciences. *Annual Review of Information Science and Technology, 45*(1), 247-294. Retrieved August 1, 2013 from the WWW: http://kalpanashankar.files.wordpress.com/2010/06/arist_data_sharing.pdf

88. Kunze, J., Cruse, P., Hu, R., Abrams, S., Hastings, K., Mitchell, C., & Schiff, L. (2011). *Practices, trends, and recommendations in technical appendix usage for selected data-intensive disciplines.* Retrieved August 1, 2013 from the WWW: http://escholarship.org/uc/item/9jw4964t#page-2

    Abstract: There is a need to establish a new publishing paradigm to cope with the deluge of data artifacts produced by data-intensive science, many of which are vital to data re-use and verification of published scientific conclusions. Due to the limitations of traditional publishing, most of these artifacts are not usually disseminated, cited, or preserved. These latent artifacts consist largely of datasets and data processing information that together form the foundations of the reasoned analyses that appear in the published literature. One promising approach to this problem of data invisibility is to wrap these artifacts in the metaphor of a "data paper", a somewhat unfamiliar bundle of scholarly output with a familiar facade. As envisioned, a data paper minimally consists of a cover sheet and a set of links to archived artifacts.

89. Lane, M. (2008). *Data citation in the electronic environment: A white paper commissioned by GBIF* [White paper]. Retrieved August 1, 2013 from the WWW: http://www.gbif.org/orc/?doc_id=4884

90. Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. Citation and peer review of data: Moving towards formal data publication. *The International Journal of Digital Curation, 6*(2), 4-37. Retrieved August 1, 2013 from the WWW: http://www.ijdc.net/index.php/ijdc/article/view/181/265

    Description: Defines publication and discusses procedures necessary to validate published data through peer review: required metadata, data refereeing, data copyright data review checklist, publication models (proxy, appendix, data archival, overlay), current data journals, citing data and existing formats, recommended citation syntax.

91. Lyon, L. (2007). Dealing with data: Roles, rights, responsibilities and relationships [Consultancy report]. Retrieved August 1, 2013 from the WWW: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

92. Lyon, L., Rusbridge, C., Neilson, C., & Whyte, A. (2010). Disciplinary approaches to sharing, curation, reuse and preservation. *DCC SCARP Final Report.* Retrieved August 1, 2013 from the WWW: http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf

93. Major, G. (2011). Impact of NASA EOS instrument data on the scientific literature: 10 years of published research results from terra, aqua, and aura. *Issues in Science and Technology Librarianship* 67. doi:10.5062/F4CC0XMJ.

    Abstract: In the absence of formal data set citation standards in the literature, there is no quantitative information on the connection between data distributed from NASA's Earth Observing System (EOS) data centers and subsequent research published using EOS data. This paper provides an analysis of a 10-year citation history of research using EOS instrument data in the peer-reviewed literature, which illustrates that the high volume of published EOS-related papers is indicative of the use of data from the NASA DAACs and comprises a significant contribution to the body of scientific knowledge about the Earth's climate.

94. Marcus, C., Ball, S., Delserone, L., Hribar, A., & Loftus, W. (2007). Understanding research behaviors, information resources, and service needs of scientists and graduate students. Retrieved from http://conservancy.umn.edu/bitstream/5546/1/Sciences_Assessment_Report_Final.pdf

    Description: Good for a general understanding of researcher behavior but little use for citation

95. Marical, L. Hemminger, B. (2010). Scientific data repositories on the web: An initial survey. *JASIST*. doi: 10.1002/asi.21339.

Abstract: Characteristics of the SDRs were explored for their role in determining groupings and for their relationship to the success of SDRs. Four of these characteristics were identified as important for further investigation: whether the SDR was supported with grants and contracts, whether support comes from multiple sponsors, what the holding size of the SDR is and whether a preservation policy exists for the SDR

96. Michner, W. Vision, T., Cruse, P. Vieglais, D., Kunze, J., & Janee, G. (2011). DataONE: Data observation network for earth — Preserving data and enabling innovation in the biological and environmental sciences. *D-Lib Magazine, 17*(½). doi:10.1045/january2011-michener

    Abstract: This paper discusses many of the issues associated with formally publishing data in academia, focusing primarily on the structures that need to be put in place for peer review and formal citation of datasets. Data publication is becoming increasingly important to the scientific community, as it will provide a mechanism for those who create data to receive academic credit for their work and will allow the conclusions arising from an analysis to be more readily verifiable, thus promoting transparency in the scientific process. Peer review of data will also provide a mechanism for ensuring the quality of datasets, and we provide suggestions on the types of activities one expects to see in the peer review of data. A simple taxonomy of data publication methodologies is presented and evaluated, and the paper concludes with a discussion of dataset granularity, transience and semantics, along with a recommended human-readable citation syntax.

97. Mooney, H, Newton, MP. (2012). The anatomy of a data citation: Discovery, reuse, and credit. *Journal of Librarianship and Scholarly Communication* 1(1):eP1035. doi: http://dx.doi.org/10.7710/2162-3309.1035

    Abstract: Citations for data must be promoted as an essential component of data publication, sharing, and reuse. Despite confounding factors, librarians and information professionals are well-positioned and should persist in advancing data citation as a normative practice across domains. Doing so promotes a value proposition for data sharing and secondary research broadly, thereby accelerating the pace of scientific research.

98. Moreau, L. (2010). The foundations for provenance on the web. *Foundations and Trends in Web Science, 2*(2-3), 99-241. doi: 10.1561/1800000010

    Abstract: As the Web allows information sharing, discovery, aggregation, filtering and flow in an unprecedented manner, it also becomes difficult to identify the original source that produced information on the Web. This survey contends that provenance can and should reliably be tracked and exploited on the Web, and investigates the necessary foundations to achieve such a vision.

99. Moreau, L. & Missier, P. (2013) PROV-DM: The PROV data model. Retrieved July 30, 2013 from the WWW: http://www.w3.org/TR/2013/REC-prov-dm-20130430/

100. Nelson, B. (2009). Data sharing: Empty archives. *Nature, 461*, 160-163. doi:10.1038/461160a

101. Page, R.D.M. (2008). Biodiversity informatics: The challenge of linking data and the role of shared identifiers. *Briefings in Bioinformatics, 9*(5), 345-54. doi: 10.1093/bib/bbn022

    Abstract: A major challenge facing biodiversity informatics is integrating data stored in widely distributed databases. Initial efforts have relied on taxonomic names as the shared identifier linking records in different databases. However, taxonomic names have limitations as identifiers, being neither stable nor globally unique, and the pace of molecular taxonomic and phylogenetic research means that a lot of information in public sequence databases is not linked to formal taxonomic names. This review explores the use of other identifiers, such as specimen codes and GenBank accession numbers, to link otherwise disconnected facts in different databases. The structure of these links can also be exploited using the PageRank algorithm to rank the results of searches on biodiversity databases. The key to rich integration is a commitment to deploy and reuse globally unique, shared identifiers [such as Digital Object Identifiers (DOIs) and Life Science Identifiers (LSIDs)], and the implementation of services that link those identifiers.

102. Parsons, M., Bruin, T., Tomlinson, S., Campbell, H., Godoy, O., LeClert, J., et al.(2009). *The state of polar data—The IPY experience.* Retrieved August 1, 2013 from the WWW: http://www.arctic.gov/downloads/Arctic_SAP/Misc/State_of_Polar_Data20100514_distribute.pdf

103. Parsons, M., Duerr, R., Minster, J. (2010). Data citation and peer review. *EOS, Transactions American Geophysical Union, 91*(34) 297-298, doi: 10.1029/2010EO340001

Abstract: A scientific publication is fundamentally an argument consisting of a set of ideas and expectations supported by observations and calculations that serve as evidence of its veracity. An argument without evidence is only a set of assertions. Scientific papers do, of course, present specific data points as evidence for their arguments, but how well do papers guide readers to the body of those data, where the data's integrity can be further examined? In practice, a chasm may lie across the path of a reviewer seeking the source data of a scientific argument.

104. Paton, N.W. (2008). Managing and sharing experimental data: Standards, tools and pitfalls. *Biochemical Society Transactions, 36*(1), 33-36. Retrieved August 1, 2013 from the WWW: http://www.mendeley.com/research/managing-and-sharing- experimental-data-standards-tools- and-pitfalls/

Abstract: The present paper discusses issues associated with the management of experimental data in the life sciences, including: the different tasks that experimental data and metadata can support, the role of standards in informing data sharing and archiving, and the development of effective databases and tools, building on these standards.

105. Peng, R. D. (2011). Reproducible research in computational science. *Science, 334* (6060), 1226-1227. doi:10.1126/science.1213847

106. Pepe, A., Mayernik, M., Borgman, C. (2010). From artifacts to aggregations: Modeling scientific life cycles on the semantic web. *Journal of the American Society for Information Science and Technology, 61*(3). doi:10.1002/asi.21263

107. Pepler, S. (2008). Citation, location and deposition in discipline and institutional repositories. *CLADDIER.* Retrieved August 1, 2013 from the WWW: http://epubs.cclrc.ac.uk/bitstream/2359/Report_II_PreservationIntentAndCompoundObjectIdentifiers-1.pdf

Description: Lots of theoretical work about citation, but no user surveys done.

108. Pepler, S. J., & O'Neil, K. (2008). *Preservation intent and collection identifiers: CLADDIER project report II.* Retrieved August 1, 2013 from the WWW: http://epubs.cclrc.ac.uk/bitstream/2359/Report_II_PreservationIntentAndCompoundObjectIdenti fiers-1.pdf

Abstract: There is ambiguity in what type of object a datasets is; with different groups of users applying different connotations. More explicit language such as "data file collection" ensures that objects are well defined. Preservation, identification and object definition are intimately linked. Using what needs to be preserved by a particular user community is an excellent way to define the boundaries and properties of datasets.

109. Peroni, S., & Shotton, D. (2012). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services and Agents on the World Wide Web, 17*, 33-34. doi:10.1016/j.websem.2012.08.001

110. Reilly, S., Schallier, W., Schrimpf, S., Smit, E., & Wilkinson, M. (2011). Report of integration of data and publications. *ODE publications*. Retrieved from http://www.alliancepermanentaccess.org/index.php/2011/10/24/ode-report-on-integration-of-data-and-publications-published/

Abstract: This report seeks to coalesce current though and opinions from numerous and diverse sources to reveal opportunities for supporting a more connected and integrated scholarly record. Four perspectives were considered, those of the Researcher, who generates or reuses primary data, Publishers, who provide the mechanisms to communicate research activities, and Libraries & Data Centers, who maintain and preserve the evidence that underpins scholarly communication and the published record.

111. Rodriguez, M., Bollen, J., & Sompel, H. (2007). A practical ontology for the large-scale modeling of scholarly artifacts and their usage. *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 278-287). New York, NY: ACM. Retrieved from http://public.lanl.gov/herbertv/papers/Papers/2007/JCDLrodriguez.pdf

Abstract: The large-scale analysis of scholarly artifact usage is constrained primarily by current practices in usage data archiving, privacy issues concerned with the dissemination of usage data, and the lack of a practical ontology for modeling the usage domain. As a remedy to the third constraint, this article presents a scholarly ontology that was engineered to represent those classes for which large-scale bibliographic and usage data exists, supports usage research, and whose instantiation is scalable to the order of 50 million articles along with their associated artifacts (e.g. authors and journals) and an accompanying 1 billion usage events. The real world instantiation of the presented abstract ontology is a semantic network model of the scholarly community which lends the scholarly process to statistical analysis and computational support. We present the ontology, discuss its instantiation, and provide some example inference rules for calculating various scholarly artifact metrics.

112. Schindler, U., Brase, J., & Diepenbroek, M. (2005). Webservices infrastructure for the registration of scientific primary data. *Research and Advanced Technology for Digital Libraries Lecture Notes in Computer Science, 3652*, 128-138. doi:10.1007/11551362_12

Description: Registration of scientific primary data, to make these data citable as a unique piece of work and not only a part of a publication, has always been an important issue. In the context of the project "Publication and Citation of Scientific Primary Data" funded by the German Research Foundation (DFG) the German National Library of Science and Technology (TIB) has become the first registration agency worldwide for scientific primary data. Registration has started for the field of earth science, but will be widened for other subjects in the future. This paper shall give an overview about the technical realization of this important usage field for a digital library.

113. Schwartz, A., Pappas, C., & Sandlow, J. (2010). Data repositories for medical education research: Issues and recommendations. *Academic Medicine, 85*(5), 837-843. doi:10.1097/ACM.0b013e3181d74562

Description: The authors explore issues surrounding digital repositories with the twofold intention of clarifying their creation, structure, content, and use, and considering the implementation of a global digital repository for medical education research data sets. The authors review digital repositories in medicine, social sciences, and education, describe the contents and scope of repositories, and present extant examples.

114. Sedransk, N., Young, L., Kelner, K., Moffitt, R., Thakar, A., Raddick, J., Ungvarsky, E., Carlson, R., Apweiler, R., Cox, L., Nolan, D., Soper, K., & Spiegelman, C. (2010). Make research data public?—Not always so simple: A dialogue for statisticians and science editors. *Statistical Science, 25*(1), 41-50. doi:10.1214/10-STS320

Description: Putting data into the public domain is not the same thing as making those data accessible for intelligent analysis. A distinguished group of editors and experts who were already engaged in one way or another with the issues inherent in making research data public came together with statisticians to initiate a dialogue about policies and practicalities of requiring published research to be accompanied by publication of the research data. This dialogue carried beyond the broad issues of the advisability, the intellectual integrity, the scientific exigencies to the relevance of these issues to statistics as a discipline and the relevance of statistics, from inference to modeling to data exploration, to science and social science policies on these issues.

115. Seeber, F. (2008). Citations in supplementary information are invisible. *Nature, 451*(7181), 887. doi:10.1038/451887d

116. Shotton, D. (2010). CiTO, the citation typing ontology. *Journal of Biomedical Semantics* 1 (Suppl. 1): S6. doi:10.1186/2041-1480-1-S1-S6

117. Shotton D (2012). The five stars of online journal articles — a framework for article evaluation. *D-Lib Magazine 18* (1/2) (January/February 2012 issue). doi:10.1045/january2012-shotton

118. Sieber, J. E., & Trumbo, B. E. (1995). (Not) giving credit where credit is due: Citation of data sets. *Science and Engineering Ethics, 1*(1), 11-20. doi:10.1007/BF02628694

Description: Adequate Citation of data sets is crucial to the encouragement of data sharing, to the integrity and cost-effectiveness of science and to easy access to the work of others. The citation behavior of social scientists who have published based on shared data was examined and found to be inconsistent with important ideals of science. Insights gained from the social sciences, where data sharing is somewhat

customary, suggest policies and incentives that would foster adequate citation by secondary users, and greater openness and sharing in other disciplines.

119. Smit, E. (2010). Abelard and Héloise: Why data and publications belong together. *D-Lib Magazine, 17*(½). doi:10.1045/january2011-smit

Description: This article explores the present state of integration between data and publications. The statistical findings are based on the project PARSE. Insight, which was carried out with the help of EU funding in 2008 - 2010. The main conclusion drawn from these findings is that currently very few conventions and best practices exist among researchers and publishers in how to handle data. There is strong preference among researchers and publishers alike for data and publications to be linked in a persistent way. To achieve that, we advocate good collaboration across the whole information chain of authors, research institutes, data centers, libraries and publishers. DataCite is an excellent example of how this might work.

120. Smit, E. (2011). Avoiding a digital dark age for data: Why publishers should care about digital preservation. *Learned publishing, 24*(1), 35-49. doi:http://dx.doi.org/10.1087/20110107

Description: This paper provides an overview of the needs and threats for digital preservation and summarizes the findings from project PARSE.Insight. This project, co-funded by the EU, contains one of the first large worldwide surveys about digital preservation including most players of the STM information chain: researchers, libraries, data managers, publishers, and research funders. One of the conclusions is that in the present data deluge, it is extremely important that all players in the information chain work together on proper digital preservation of all research output, to ensure its future usability, understandability, and authenticity. This poses a new role for publishers who can ensure better discoverability and citability via good linking and integration of data and publications.

121. Starr, J., & Gastl, A. (2011). isCitedBy: A metadata scheme for DataCite. *D-Lib Magazine, 17*(1/2). doi:10.1045/january2011-starr

Description: The DataCite Metadata Scheme is being designed to support dataset citation and discovery. It features a small set of mandatory properties, and an additional set of optional properties for more detailed description. Among these is a powerful mechanism for describing relationships between the registered dataset and other objects. The scheme is supported organizationally and will allow for community input on an ongoing basis.

122. Stodden, V. (2009). The legal framework for reproducible scientific research: Licensing and copyright. *Computing in Science and Engineering, 11*(1), 35-40. doi:http://dx.doi.org/10.1109/MCSE.2009.19

123. Takeda, K. (2010). *Institutional data management blueprint project*. Retrieved August 1, 2013 from the WWW: http://www.southamptondata.org/uploads/7/3/0/0/730051/idmbinitialfindingsreportv4.pdf.

Talk here: August 1, 2013 from the WWW: http://www.eduserv.org.uk/newsandevents/events/eduserv-symposium-2011/the-institutional-data-management-blueprint-project

Slides here: August 1, 2013 from the WWW: http://www.slideshare.net/eduserv/institutional-data-management-blueprint-7979017

124. Takeda, K., Brown, M., Coles, S., Carr, L., Earl, G., Frey, J., Hancock, P., White, W., Nichols, F., Whitton, M., Gibbs, H., Fowler, C., Wake, P., & Patterson, S. (2010). Data management for all: The institutional data management blueprint project. *6th International Digital Curation Conference*. Retrieved from http://eprints.soton.ac.uk/169533/1/6th_international_digital_curation_conference__idmb_final_paper_revised.pdf

Description: This paper describes the Institutional Data Management Blueprint (IDMB) project, which aims to create a practical and attainable institutional framework for managing research data throughout its lifecycle that facilitates ambitious national and international e-research practice. The objective is to produce a framework for managing research data across the whole lifecycle that encompasses a whole institution (exemplified by the University of Southampton) and based on an analysis of current data management requirements for a representative group of disciplines with a range of different data.

125. Thessen, A., & Patterson, D. (2011). *Data issues in the life sciences* [*White paper*]. Retrieved August 1, 2013 from the WWW: http://dataconservancy.org/sites/default/files/Data%20Issues%20in%20the%20Life%20Sciences%20White%20Paper.pdf

126. Tilmes, C., Yesha, Y., & Halem, M. (2011). Distinguishing provenance equivalence of earth science data. *Procedia Computer Science, 4*, 548–557. doi:http://dx.doi.org/10.1016/j.procs.2011.04.057

Description: This paper discusses scientific equivalence and essential provenance for scientific reproducibility. We use the example of an operational earth science data processing system to illustrate the application of the technique of cascading digital signatures or "hash chains" to precisely identify sets of granules and as provenance equivalence identifiers to distinguish data made in an equivalent manner

127. *Toward a Consistent Policy for Reporting Geochemical Data in Publications and to Databases.* (2008). Policy adopted by the Editors' Roundtable at the Goldschmidt Conference. Retrieved August 1, 2013 from the WWW: http://www.geoinfogeochem.org/sites/geoinfogeochem.org/files/Policy_GeochemDataPubl_v1.1_0.pdf

128. Urban, E., Leadbetter, A., Moncoiffe, G., Pissierssens, P., Raymond, L., & Pikula, L. (2012). Pilot projects for publishing and citing ocean data. Eos, Transactions American Geophysical Union. 93(43):425-426. doi: 10.1029/2012EO430001

129. Van der Graaf, M. and Waaijers, L. (2011). A Surfboard for riding the wave: Towards a four country action programme on research data. A Knowledge Exchange Report. Retrieved August 1, 2013 from the WWW: www.knowledge-exchange.info/surfboard

Abstract: The Riding the Wave report calls for a collaborative data infrastructure that will enable researchers and other stakeholders from education, society and business to use, re-use and exploit research data to the maximum benefit of science and society. The Knowledge Exchange (KE) partners have embraced this vision. This paper presents an overview of the present situation with regard to research data in Denmark, Germany, the Netherlands and the United Kingdom and offers broad outlines for a possible action programme for the four countries in realizing the envisaged collaborative data infrastructure. An action programme at the level of four countries needs the involvement of all stakeholders from the scientific community. We identified four key drivers: incentives and training in relation to researchers in their role as data producers and users of information infrastructures infrastructure and funding of the infrastructure in relation to further developments in data logistics.

130. Vision, T.J. (2010). Open data and the social contract of scientific publishing. *American Institute of Biological Sciences, 60*(5), 330-331. doi:http://dx.doi.org/10.1525/bio.2010.60.5.2

131. W3C. Incubator report (2010). Retrieved August 1, 2013 from the WWW: http://www.w3.org/2005/Incubator/prov/XGR-prov-20101214/.

Abstract: Given the increased interest in provenance in the Semantic Web area and in the Web community at large, the W3C established the Provenance Incubator Group as part of the W3C Incubator Activity with a charter to provide a state-of-the art understanding and develop a roadmap in the area of provenance and possible recommendations for standardization efforts. This document summarizes the findings of the group. Slides here; http://www.w3.org/2005/Incubator/prov/wiki/File:Provenance-XG-Overview.pdf

132. Walker, D., (2010). The physics of complex systems in information and biology. Retrieved August 1, 2013 from the WWW: http://gradworks.umi.com/33/86/3386263.html

Abstract: Citation networks have re-emerged as a topic intense interest in the complex networks community with the recent availability of large-scale data sets. The ranking of citation networks is a necessary practice as a means to improve information navigability and search. Unlike many information networks, the aging characteristics of citation networks require the development of new ranking methods. To account for strong aging characteristics of citation networks, we modify the PageRank algorithm by initially distributing random surfers exponentially with age, in favor of more recent publications. The output of this algorithm, which we call CiteRank, is interpreted as approximate traffic to individual publications in a simple model of how researchers find new information. We optimize parameters of our algorithm to achieve the best performance. The results are compared for two rather different citation

networks: all American Physical Society publications between 1893-2003, and the set of high-energy physics theory (hep-th) preprints. Despite major differences between these two networks, we find that their optimal parameters for the CiteRank algorithm are remarkably similar. The advantages and performance of CiteRank over more conventional methods of ranking publications are discussed.

133. Wallis, J., Borgman, C., Mayernik, M. & Pepe, A. (2008). Moving archival practices upstream: An exploration of the life cycle of ecological sensing data in collaborative field research. *International Journal of Digital Curation, 3*(1), 114-126. Retrieved August 1, 2013 from the WWW: http://www.ijdc.net/index.php/ijdc/article/viewFile/67/46

    Abstract: The success of eScience research depends not only upon effective collaboration between scientists and technologists but also upon the active involvement of data archivists. Archivists rarely receive scientific data until findings are published, by which time important information about their origins, context, and provenance may be lost. Research reported here addresses the life cycle of data from collaborative ecological research with embedded networked sensing technologies. A better understanding of these processes will enable archivists to participate in earlier stages of the life cycle and to improve curation of these types of scientific data. Evidence from our interview study and field research yields a nine-stage life cycle. Among the findings are the cumulative effect of decisions made at each stage of the life cycle; the balance of decision-making between scientific and technology research partners; and the loss of certain types of data that may be essential to later interpretation.

134. Waters, D. (2004). Building on success, forging new ground: The question of sustainability. *First Monday, 9*(5). doi:http://dx.doi.org/10.5210%2Ffm.v9i5.1148

    This paper focuses on three factors that contribute to the sustainability of digital scholarly resources. First, the development of such resources depends on a clear definition of the audience and the needs of users. Second, the resource must be designed to take advantage of economies of scale. Third, to create an enduring resource, careful attention is needed to the design of the organization that will manage the resource over time.

135. Wynholds, L. (2011). Linking to scientific data: Identity problems of unruly and poorly bounded digital objects. *International Journal of Digital Curati*on, *6*(1), 214-225. doi:10.2218/ijdc.v6i1.183

    Abstract: This paper explores some of the ways in which scientific data is an unruly and poorly bounded object, and goes on to propose that in order for datasets to fulfill the roles expected for them, the following identity functions are essential for scholarly publications: (i) the dataset is constructed as a semantically and logically concrete object, (ii) the identity of the dataset is embedded, inherent and/or inseparable, (iii) the identity embodies a framework of authorship, rights and limitations, and (iv) the identity translates into an actionable mechanism for retrieval or reference.

136. Yanhua, Z., and Lianglin, H. (2012). Analyzing the Influence of Scientific Database Based on the Third-party Cited Marks. *China Science & Technology Resources Review*, *44*, 6, 17-22. Retrieved August 1, 2013 from the WWW: http://www.zgkjzydk.com.cn/ch/reader/view_abstract.aspx?file_no=20120604&flag=1

137. Zettsu, K., Gonzales, E., Ong, B. T., Murayama Y. (2012, October). Cross-Database Search for Interdisciplinary Use of Large-Scale, Multi-Domain and Heterogeneous Databases. Paper presented at the 23rd International CODATA Conference, Taipei, China.

Posters, Charts

1. Enriquez, V., Judson, S.W., Weber, N.M., Allard, S., Cook, R.B., Piwowar, H.A., Sandusky, R.J.,Vision, T.J., & Wilson, B. (2010). Data citation in the wild. Chicago, IL: IDCC. Retrieved August 1, 2013 from the WWW: http://www.readcube.com/articles/10.1038/npre.2010.5452.1?locale=en

2. Newton, M. Mooney, H, & Witt, M. A description of data citation instructions in style guides. Retrieved August 1, 2013 from the WWW: http://docs.lib.purdue.edu/lib_research/121/

3. NIH data sharing regulations/policy/guidance/ chart for NIH awards. Retrieved August 1, 2013 from the WWW:grants.nih.gov/grants/policy/data_sharing/data_sharing_chart.doc

4. Pepe, A., Borgman, C. (nd) Integrating scientific data into scholarly value chains. Retrieved August 1, 2013 from the WWW: http://research.microsoft.com/en-us/events/ersymposium2009/integrating_scien_data_scholarly_value_chains.pdf

5. Piwowar,H. Chapman, W. (2007) Examining the uses of shared data, Poster. Retrieved August 1, 2013 from the WWW:http://precedings.nature.com/documents/425/version/3/files/npre2007425-3.pdf

Presentations, PowerPoints, Videos

1. Brase, J. (2012). DataCite revisited: Citing data in the 21st century, at long last. Retrieved August 1, 2013 from the WWW: http://www.youtube.com/watch?v=bN_AqbI-hmo

2. Callaghan, S. (2011). Making data a first class scientific output: data citation and publication by NERC's environmental center. Retrieved August 1, 2013 from the WWW: http://vimeo.com/34338054.

3. Chavan, V. (2010). Data Citation Mechanism and Services for primary biodiversity data. Global Biodiversity Information Facility. Retrieved August 1, 2013 from the WWW: http://www.google.com/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=9&ved=0CGgQFjAI&url=http%3A%2F%2Fwww.tdwg.org%2Ffileadmin%2F2010conference%2Fslides%2FChavan_DataCitation.ppt&ei=peZkT6K1G4GvsgK6guy2Dw&usg=AFQjCNECBPHpEG_AbaNvIyIsfaVofXebiA&sig2=AmYwI-nGlIJ5AjpFFy4RfA

4. Chen, R. S. and Downs, R. R. (2010). Evaluating the use and impacts of scientific data. National Federation of Advanced Information Services (NFAIS) Workshop, Assessing the Usage and Value of Scholarly and Scientific Output: An Overview of Traditional and Emerging Approaches. Philadelphia, PA, November 10, 2010. Retrieved August 1, 2013 from the WWW: http://info.nfais.org/info/ChenDownsNov10.pdf

5. Coggins, J. (2009, Feb). A researcher's perspective: the value and challenge of data. A National Research Data Service for the UK? An international conference on the UK Research Data Service feasibility study, London, United Kingdom. Retrieved August 1, 2013 from the WWW: http://www.ukoln.ac.uk/events/ukrds-2009/programme/

6. Edmunds, S. Data dissemination, difficulties, data citation, DOIs (and Giga Science). Retrieved August 1, 2013 from the WWW: http://www.youtube.com/watch?v=AlYFa83aCWA

7. ESDS International. Citing data. Retrieved August 1, 2013 from the WWW: http://youtu.be/NDrNHRjtd4g

8. IASSIST. (2011, June). Session B2: The IASSIST SIGDC Presents: Perspectives on data citation; Session C2: DataCite - Making Data Citable. Retrieved August 1, 2013 from the WWW: http://www.iassistdata.org/conferences/archive/2011

9. King, G. (2007). An introduction to the Dataverse Network as an infrastructure for data sharing. Retrieved August 1, 2013 from the WWW: http://www.youtube.com/watch?v=fgn6dmfsZ_M

10. Lautenschlager, M., Hock, H., Brase, J. & Waszkewitz, S. (2008, March). Publication & citation of primary data at WDC Climate (WDCC). MPG eScience seminar: Persistent identifier. Garching, Germany. Retrieved August 1, 2013 from the WWW: http://colab.mpdl.mpg.de/mediawiki/images/3/30/ESci08_Sem_1_Primary_data_registration_Lautenschlager.pdf

11. Linares, F. A. (2008). Metadata & scientific data: Integrating DDE, STTR, and ICSTI initiatives. Information International Associates, Inc. Presentation to CENDI Federal STI Managers Group. Retrieved from http://www.cendi.gov/presentations/10-29-08_CENDI_Azpurua.pdf

12. McMahon, B. (2010). Integrating data with publications: Greater interactivity and challenges for long-Term preservation of the scientific record. Crystallography Journals Online. Retrieved August 1, 2013 from the WWW: http://www.codata.org/10Conf/abstracts-presentations/Sessions%20F/F1/F1-McMahon.pdf

13. MIT Library. A day on the life of a dataset. Retrieved August 1, 2013 from the WWW: http://libraries.mit.edu/guides/subjects/data-management/Managing%20Research%20Data%20101.pdf

14. Mooney, H. (2010, June). Data reference in depth: Citation. IASSIST 2010 Conference. Retrieved August 1, 2013 from the WWW: http://www.iassistdata.org/downloads/2010/2010_g1_mooney.pdf

15. Piwowar, H. (2011). 7 data citations challenges illustrated with examples (includes elephants). JISC Managing Research data. Retrieved August 1, 2013 from the WWW:http://www.slideshare.net/hpiwowar/7-data-citation-challenges-illustrated-with-data-includes-elephants

16. RiverValleyTV. E-books and e-content, 2010. University College of London. Retrieved August 1, 2013 from the WWW: http://river-valley.tv/conferences/ebooks-econtent-2010

Reports (Government/ Committee)

1. ANDS Technical Working Group. (2007). *Towards the Australian data commons*. Retrieved August 1, 2013 from the WWW: ands.org.au/towardstheaustraliandatacommons.pdf

2. Atkins, D., Droegemeier, K., Feldman, S., Garcia-Molina, H., Klein, M., Messerschmitt, D., Wright, M. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure 2003-01*. Retrieved August 1, 2013 from the WWW: http://www.nsf.gov/cise/sci/reports/atkins.pdf

3. Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (2010). *Sustainable economics for a digital planet: Ensuring long-term access to digital information*. Retrieved August 1, 2013 from the WWW:http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

4. Bohn, R., & Short, J. (2009). *How much information? Global Information Industry Center Report on American Consumers*. Retrieved August 1, 2013 from the WWW: http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf

   Description: Focusing on information consumption.

5. Boulton, G., Campbell, P., Collins, B., Elias, P., Hall, W., Laurie, G., O'Neill, O. . . .Walport, M. (2012). *Science as an open enterprise: The Royal Society Science Policy Centre report.* Retrieved July 30, 2013 from the WWW: http://royalsociety.org/policy/projects/science-public-enterprise/report/

   Description: Exploring massive amounts of data using modern digital technologies has enormous potential for science and its application in public policy and business. The report maps out the changes that are required by scientists, their institutions and those that fund and support science if this potential is to be realised.

6. CCSDS (Consultative Committee for Space Data Systems). (2002*). Reference model for an open archival information system*. (OAIS). CCSDS 650.0-B-1, 1. Washington, DC: CCSDS Secretariat. Retrieved August 1, 2013 from the WWW: http://public.ccsds.org/publications/archive/650x0m2.pdf

7. CENDI. (2004). *Persistent identification: A key component of an e-government infrastructure*. CENDI Persistent Identification Task Group. http://www.cendi.gov/publications/04-2persist_id.html

8. Chavan, V. (2012b). *Recommended practices for citation of data published through the GBIF network*. Copenhagen: GBIF Secretariat. Retrieved July 30, 2013 from the WWW: http://links.gbif.org/gbif_best_practice_data_citation_en_v1

   Description: The GBIF Data Publishing Framework Task Group established in 2009, recommended that GBIF institutionalize a 'data citation mechanism' and establish a 'data citation service' facilitating deep data citation, and registration and resolving of citations (Moritz et.al, 2011). As an early uptake of this recommendation, GBIF in consultation with a group of experts has come up with recommended practices for citing biodiversity data. This document recommends a set of styles for (a) Publisher-based citations, and (b) Query-based citations. The recommended sets of styles for publisher-based citations are for immediate uptake by data publishers, data owners, data custodians, and data aggregators.

9. CSIR. (2011). *Geoportal offers larger spatial data seta at the click of a button*. Ecosystems. Earth observation. Retrieved August 1, 2013 from the WWW: http://www.csir.co.za/nre/ecosystems/Geoportal.html

   Description: A briefing of the history and responsibilities of the CSIR, and makes available the Geospatial information through the Geo-portal. The improvements, developments and strategies of the CSIR in are all examples of what is explained.

10. Finch, J. (2012). *Accessibility, sustainability, excellence: How to expand access to research publications: Report of the working group on expanding access to published research findings*. Retrieved July 30, 2013 from the WWW: http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf

11. High Level Expert Group on Scientific Data. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. Retrieved August 1, 2013 from the WWW: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

    Description: Support for scientific e-infrastructure that supports seamless access, use, re-use, and trust of data.

12. Interagency Working Group on Digital Data. (2009). *Harnessing the power of digital data for science and society: Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council.* Retrieved August 1, 2013 from the WWW: http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

13. International Association of Scientific, Technical, and Medical Publishers. (2007). *Brussels Declaration of 2007*. Retrieved August 1, 2013 from the WWW: http://www.stm-assoc.org/brussels-declaration/

14. Joint Task Force on Library Support for E-Science (2007). *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*. Retrieved August 1, 2013 from the WWW: http://www.arl.org/bm~doc/ARL_EScience_final.pdf

    Description: Article relates to eScience and libraries. Citation was not mentioned.

15. Kotarski, R., Reilly, S., Schrimpf, S.,Smit, E., Walshe, K. (2012, June21). Report on best practices for citability of data and on evolving roles in scholarly communication. Retrieved July 30, 2013 from the WWW: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/08/ODE-ReportBestPracticesCitabilityDataEvolvingRolesScholarlyCommunication.pdf

16. Mayernik, M., 2013: *Bridging data lifecycles: Tracking data use via data citations workshop report*. NCAR Technical Note NCAR/TN-494+PROC. doi: 10.5065/D6PZ56TX.

17. Murphy, C. (1982). *Micrometeorological data for the energy balance and the exchange of carbon dioxide between a forest and the atmosphere* (Report No. AC09-76SR00001). Washington, D.C.: Department of Commerce. Retrieved August 1, 2013 from the WWW: http://www.ntis.gov/search/product.aspx?ABBR=DE82019300

    Description: The data reported was collected to measure the energy balance and carbon dioxide flux of a young pine plantation. The data set consists of half-hour averages of the meteorological parameters.

18. National Digital Stewardship Alliance. (2011). Response to office of science and technology policy request for information on public access to digital data resulting from federally funded scientific research. Retrieved August 1, 2013 from the WWW:: http://digitalpreservation.gov/documents/NDSA_ResponseToOSTP.pdf

19. National Information Standards Organization. (2013). *Recommended practices for online supplemental journal article materials*. (Report No. NISO RP-15-201x). Retrieved July 30, 2013 from the WWW: http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf

20. National Science Foundation (2007). *Cyberinfrastructure vision for 21st century discovery*. Arlington, VA: National Science Foundation. Retrieved August 1, 2013 from the WWW: http://www.nsf.gov/pubs/2007/nsf0728/

21. National Science Foundation. (2011). *Advisory Committee for Cyberinfrastructure, and Task Force on Data and Visualization: Final Report.* Arlington, VA: National Science Foundation. Retrieved August 1, 2013 from the WWW: http://www.nsf.gov/od/oci/taskforces/TaskForceReport_Data.pdf

22. National Science Foundation. (2011). *Digital research data sharing and management*. Arlington, VA: National Science Foundation. Retrieved August 1, 2013 from the WWW: http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf \

Description: Recommendations are organized under four areas: commitment to sharing; reproducibility; education, training, and workforce development; and longevity and sustainability.

23. National Science Foundation. (2011). *Division of Ocean Sciences Sample and Data Policy*. Arlington, VA: National Science Foundation. Retrieved August 1, 2013 from the WWW: http://www.nsf.gov/pubs/2011/nsf11060/nsf11060.pdf

24. Organization for Co-operation and Development. (2007). *OECD principles and guidelines for access to research data from public funding*. Retrieved August 1, 2013 from the WWW: http://www.oecd.org/dataoecd/9/61/38500813.pdf

25. Research Information Network. (2008). *To share or not to share*. Retrieved August 1, 2013 from the WWW: http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs

26. UNESCO (2008). *SCOR/IODE workshop on data publishing*. IOC Workshop (Report No. 207). Retrieved August 1, 2013 from the WWW: http://www.scor-int.org/Publications/wr207.pdf

27. World Economic Forum. (2012). *Big data, big impact: New possibilities for economic development.* Retrieved July 30, 2013 from the WWW: http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf

Standards

1. ANSI/NISO Z39.29. (2005 R2010). Bibliographic references. Bethesda, MD: NISO Press. Retrieved August 1, 2013 from the WWW: http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key:ustring:iso-8859-1=87775a75d6ea19921a41d75b2fb012b0d6339b3a

2. DublinCore. Retrieved August 1, 2013 from the WWW: http://dublincore.org/

3. IEEE. Retrieved August 1, 2013 from the WWW: http://standards.ieee.org/

4. ISO 690:2010. Information and documentation: Guidelines for bibliographic references and citations to information resources. Retrieved August 1, 2013 from the WWW: http://www.iso.org/iso/catalogue_detail.htm?csnumber=43320

5. ISO 14721:2012. Space Data and Information Transfer Systems—A Reference Model for An Open Archival Information System (OAIS). International Organization for Standardization. Retrieved August 1, 2013 from the WWW: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683

6. ISO 19110:2005, Geographical Information- Methodology for Feature Cataloguing Retrieved August 1, 2013 from the WWW: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39965

Surveys & Studies

1. Banach, M., Li, Y. (2011). Institutional repositories and digital preservation: Assessing current practices at research libraries. *D-Lib Magazine, 17*(5/6). Retrieved August 1, 2013 from the WWW: http://works.bepress.com/meghan_banach/5/

Description: In spring 2010, authors from the University of Massachusetts Amherst conducted a national survey on digital preservation of Institutional Repository (IR) materials among Association of Research Libraries (ARL) member institutions. Examining the current practices of digital preservation of IR materials, the survey of 72 research libraries reveals the challenges and opportunities of implementing digital preservation for IRs in a complex environment with rapidly evolving technology, practices, and standards. Findings from this survey will inform libraries about the current state of digital preservation for IRs.

2. Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). Keeping research data safe: A cost model and guidance for UK Universities. Joint Information Systems Committee. Retrieved August 1, 2013 from the WWW: http://sitecore.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf

   Description: Addresses costs associated with managing data.

3. Beagrie, N., Lavoie, B., & Woollard, M. (2010). Keeping research data safe 2. Joint Information Systems Committee Retrieved August 1, 2013 from the WWW: http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf

4. Beagrie, N., Beagrie, R., & Rowlands, I. (2009). Research data preservation and access: The views of researchers. Ariadne 60. Retrieved August 1, 2013 from the WWW: http://www.ariadne.ac.uk/issue60/beagrie-et-al/

   Description: Findings from a UKRDS survey of researchers' views on and practices for preservation and dissemination of research data in four UK universities (Bristol, Leeds, Leicester, and Oxford) and place them in the wider UK and international context.

5. DCC SCARP Synthesis Project. (2010). Data dimensions: Disciplinary differences in research data sharing, reuse and long term viability: A comparative review based on sixteen case studies. Retrieved August 1, 2013 from the WWW: http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP-Synthesis.pdf

6. EDUCAUSE (2006). IT engagement in research: A baseline study. Retrieved August 1, 2013 from the WWW: http://www.educause.edu/ECAR/ITEngagementinResearchABaselin/158595

   Description: This ECAR study explores the practices and perspectives of IT organizations that support the academic research enterprise. To collect, analyze, and distribute information across an expanding range academic disciplines and geographic locations, research efforts rely heavily on IT infrastructure, people, and a broad range of IT services. Ever-larger data sets are being collected and shared, simulations and visualization are becoming routine tools, and the co-evolution of science and computing increasingly requires scientists to have solid grounding in information management. This study reports the results of a variety of research initiatives: a literature review, quantitative and qualitative data from 328 higher education institutions (315 U.S. and 13 Canadian institutions), and five in-depth cases studies. In addition, ECAR published What Do Researchers Need? Higher Education IT from the Researcher's Perspective, to supplement this study. (Citation not discussed)

7. Gibbs, H. (2009). Southampton data survey: Our experience and lessons learned. Retrieved August 1, 2013 from the WWW: http://ie-repository.jisc.ac.uk/304/

8. Harley, D., Acord, S., Earl-Novell, S., Lawrence, S., & King, C. (2010). Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines. Berkeley, CA: eScholarship. Retrieved August 1, 2013 from the WWW: http://escholarship.org/uc/cshe_fsc

   Description: This report brings together the responses of 160 interviewees across 45, mostly elite, research institutions to closely examine scholarly needs and values in seven selected academic fields: archaeology, astrophysics, biology, economics, history, music, and political science.

9. Johnston, L. (2010). User-needs assessment of the research cyberinfrastructure for the 21st century. Paper presented at 31st Annual IATUL Conference, West Lafayette, IN. Retrieved August 1, 2013 from the WWW: http://docs.lib.purdue.edu/iatul2010/conf/day1/5/

   Abstract: In 2009 our team conducted an extensive user-needs assessment of 780 university faculty, research staff, and graduate students. The PEL survey assessed the current and future cyberinfrastructure needs in the following areas: data storage, data management, and networking infrastructure; collaboration with other researchers; tools and applications; high performance computing; and learning and workforce development. Citation not addressed

10. Lyman, V. (2003). How Much Information? 2003 Retrieved August 1, 2013 from the WWW: http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf

    Description: Focuses on information production.

11. Maron, N. L., & Smith, Kirby (2008). Current models of digital scholarly communication: Results of an investigation conducted by Ithaka Strategic Services for the Association of Research Libraries. *The Journal of Electronic Publishing, 12*(1). doi:http://dx.doi.org/10.3998/3336451.0012.105

12. Mooney, H. (2011). Citing data sources in the social sciences: Do authors do it? *Learned Publishing, 24*(2),99-108. doi:10.1087/20110204. Retrieved August 1, 2013 from the WWW: http://staff.lib.msu.edu/mooneyh/myresearch/HMooney_Citingdatasources_preprint.pdf

    Abstract: It is expected that authors will provide citations for all papers referenced in their writings. The necessity of providing citations for data is not so widely recognized. Proponents of the data sharing movement have advocated for the citation of datasets in order to recognize contributions and enhance access. This study examines a sample of papers from the Inter-University Consortium for Political and Social Research (ICPSR) Bibliography of Data-Related Literature that are based on secondary analysis of datasets available in the ICPSR data archive to determine the data citation practices of authors. The results indicate that many authors fail to cite the data used in secondary analysis studies. Possible reasons for the dismal state of data citation practices are considered including the recent introduction of data into the scholarly record and its marginalization as an information format. Updating citation practices to include datasets will support data sharing and foster responsible scholarship.

13. Pienta, A., Alter, G., & Lyle, J. (2010). The enduring value of social science research: The use and reuse of primary research data. Paper presented at *The Organisation, Economics and Policy of Scientific Research* workshop, Torino, Italy. Retrieved August 1, 2013 from the WWW: http://deepblue.lib.umich.edu/handle/2027.42/78307

    Description: Multivariate models of count of publications suggest that data sharing, especially sharing data through an archive, leads to many more times the publications than not sharing data.

14. Piwowar, H., & Chapman, W. (2008). A review of the journal policies for sharing research data. *AMIA Annual Symposium Proceedings 2008* (pp. 596-600). Published online. Retrieved August 1, 2013 from the WWW: http://ocs.library.utoronto.ca//index.php/Elpub/2008/paper/view/684

    Abstract: The purpose of this study is to understand the current state of data sharing policies within journals, the features of journals which are associated with the strength of their data sharing policies, and whether the strength of data sharing policies impact the observed prevalence of data sharing. Of the 70 journal policies, 18 (26%) made no mention of data sharing requirements within their Instruction to Author statements. Another 11 policies (16%) included requests or requirements for sharing other types of data (usually DNA and protein sequences), but no statement covering data in general or microarray data in particular. Of the 42 journals (60%) with a data sharing policy applicable to microarrays, 24 (34% of 70) had a general statement about data sharing and 38 (54% of 70) covered microarrays explicitly. We classified 18 (26% of 70) of these policies as moderate and 24 (34% of 70) of the policies as strong. Data sharing policy was associated with impact factor.

15. Piwowar, H., & Chapman, W. (2010). Public sharing of research datasets: A pilot study of associations. *Journal of Informetrics, 4*(2), 148-156. Retrieved August 1, 2013 from the WWW:http://www.sois.uwm.edu/MetricsPreCon/documentation/Piwowar_Chapman_Sharing.pdf

    Abstract: In this pilot study, we analyze the association between the frequency with which study investigators share their gene expression microarray data in public databases and whether the study is subject to the NIH data sharing plan requirements, journal data sharing requirements, journal impact factor, and investigator experience. Across 397 recent microarray studies, we find that investigators are more likely to share their raw dataset when their study is published in a high impact journal, when their study is published in a journal with an enforceable data-sharing requirement, and when the first and/or last authors have higher levels of career experience and impact.

16. Piwowar, H., Day, R. & Fridsma, D. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE, 2*(3), e308. doi:10.1371/journal.pone.0000308

    Abstract: We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p = 0.006$)

associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

17. Polydoratou, P. (2007). Use of digital repositories by chemistry researchers: Results of a survey. Program: *Electronic Library and Information Systems, 41*, 386–399. doi:10.1108/00330330710831594

    Abstract: This paper aims to present findings from a survey that aimed to identify the issues around the use and linkage of source and output repositories and the chemistry researchers' expectations about their use. This survey was performed by means of an online questionnaire and structured interviews with academic and research staff in the field of chemistry. A total of 38 people took part in the online questionnaire survey and 17 participated in face-to-face interviews, accounting for 55 responses in total. Members of academic and research staff in chemistry from institutions in the UK were, in general, favorably disposed towards the idea of linking research data and published research outputs, believing that this facility would be either a significant advantage or useful for the research conducted in the domain. Further information about the nature of the research that they conduct, the type of data that they produce, the sharing and availability of research data and the use and expectations of source and output repositories is also discussed. Research limitations/implications – Interpretation of the results must recognize that the majority of the interviewees worked in the area of theoretical/computational chemistry and therefore their views may not be representative of other chemistry research fields.

18. Randall, R., Smith, J., Clark, K. & Foster, N. (2009). The next generation of academics: A report on a study conducted at the University of Rochester. Retrieved August 1, 2013 from the WWW: http://hdl.handle.net/1802/6053

    Abstract: This document reports on the user research portion of "Enhancing Repositories for the Next Generation of Academics" (IMLS Grant No. LG-06-06-0051). We conducted user research from December 2006 through March 2008 to support development of a suite of authoring tools to be integrated into an institutional repository. Our understanding of the work practices of graduate students enabled us to design the authoring tools to meet their needs for individual and collaborative writing and to make it easy for them to move completed documents from the authoring system into the repository.

19. Research Information Network. (2011). Data centers: Their use, value, and impact. Joint Information Systems Committee. Retrieved August 1, 2013 from the WWW: http://www.rin.ac.uk/our-work/data-management-and-curation/benefits-research-data-centres

20. Research Information Network. (2011). Information handling in collaborative research: An exploration of five case studies. Joint Information Systems Committee. Retrieved August 1, 2013 from the WWW: http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/collaborative-research-case-studies

    Abstract: The case studies focus on the behaviors and needs of researchers working on both sides of collaborations between higher education institutions and an external partner. The overall aim of the case studies was to: understand how researchers manage the discovery, access, use, creation, sharing and dissemination of Information resources, within the research project and with external partners; provide comparisons between the behaviors and needs of researchers in different types of collaborations; identify barriers to more effective use of information in collaborations, and provide recommendations on how such barriers might be overcome.

21. Research Information Network. (2011). Physical sciences case studies: Information use and discovery. Joint Information Systems Committee. Retrieved August 1, 2013 from the WWW: http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/physical-sciences-case-studies-use-and-discovery-

    Abstract: This project focused on the behaviors and needs of researchers working in a number of subject and disciplinary areas in the physical sciences. It follows the previous rounds of case studies in the life sciences and the humanities. The report finds that information practices in the physical sciences are highly discipline-specific. New technologies are only adopted if they make life noticeably better:

researchers will not change from their habitual behaviors if they cannot see any advantage in doing so. There is a particularly noticeable difference between the complex approaches to computation in many disciplines, and the simple approaches to information management.

22. Research Information Network. (2011). Reinventing research? Information practices in the humanities. Joint Information Systems Committee. Retrieved August 1, 2013 from the WWW: http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities

    Description: This project focuses on the behaviors and needs of researchers working in a number of subject or disciplinary areas in the humanities. The authors follow the first round of case studies in the life sciences.

23. Research Information Network (2011). The value of libraries for research and researchers. Joint Information Systems Committee. Retrieved August 1, 2013 from the WWW: http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/value-libraries-research-and-researchers

    Abstract: This jointly commissioned RIN and RLUK report presents the findings of a systematic study of the value of the services that libraries in the UK provide to researchers, and of the contributions that libraries from a wide range of institutions make to institutional research performance. The aim was to identify the key characteristics of library provision to support research in successful UK universities and departments.

24. Shaon, A., & Woolf, A. (2010). Long-term preservation for INSPIRE: A metadata framework and geo-portal implementation. D-Lib Magazine, 17(9/10). Retrieved August 1, 2013 from the WWW: inspire.jrc.ec.europa.eu/events/conferences/inspire_2010/abstracts/55.doc

    Description: The article discusses the pilot study that was done. The study has twofold aims: the investigation of feasibility of the long-term preservation of research databases and to test the technical platform for the generic model of the Open Archival Information System (OAIS). It also discuss three databases that were tested in the study.

25. Simmhan, Y., Plale, B., & Gannon, D. (2005). A survey of data provenance in e-science. *ACM SIGMOD, 34*(3). Retrieved August 1, 2013 from the WWW: http://www.sigmod.org/publications/sigmod-record/0509/p31-special-sw-section-5.pdf

    Abstract: In this paper we create a taxonomy of data provenance characteristics and apply it to current research efforts in e-science, focusing primarily on scientific workflow approaches. The main aspect of our taxonomy categorizes provenance systems based on why they record provenance, what they describe, how they represent and store provenance, and ways to disseminate it. The survey culminates with an identification of open research problems in the field.

26. Soehner, C., Steeves, C., & Ward, J. (2010). e-Science and data support services: A survey of ARL members. Paper presented at 31st Annual IATUL Conference, West Lafayette, IN. Retrieved August 1, 2013 from the WWW:http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1015&context=iatul2010

    Description: Surveyed users, but didn't ask about citation.

27. Sukovic, S. (2009). References to e-texts in academic publications. Journal of Documentation, 65(6), 997-1015. doi:10.1108/00220410910998960

    Abstract: The purpose of this paper is to explore roles of electronic texts (e-texts) in research enquiry in literary and historical studies, and to deepen the understanding of the nature of scholars' engagement with e-texts as primary materials. The study includes an investigation of references to e-texts and discussions about researchers' citation practices in interviews. Qualitative methodology was used to explore scholars' interactions with e-texts in 30 research projects. A combination of quantitative and qualitative methods was used to examine citations and any other acknowledgments of e-texts in participants' prepublications and published works. In-depth semi-structured interviews provided data for findings about researchers' citation practices. Formal acknowledgments of e-texts do not represent the depth and breadth of researchers' interactions with e-texts. Assessments of the relevance and trustworthiness of e-texts, as well as considerations of disciplinary cultures, had some impact on

researchers' citation practices. The findings have implications for the development of standards and institutional support for research in the humanities.

28. Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., Manoff, M., Frame, M., & Neylon, C. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE, 6*(6). doi:10.1371/journal.pone.0021101

    Abstract: A total of 1329 scientists participated in this survey exploring current data sharing practices and perceptions of the barriers and enablers of data sharing. Scientists do not make their data electronically available to others for various reasons, including insufficient time and lack of funding. Most respondents are satisfied with their current processes for the initial and short-term parts of the data or research lifecycle (collecting their research data; searching for, describing or cataloging, analyzing, and short-term storage of their data) but are not satisfied with long-term data preservation. Many organizations do not provide support to their researchers for data management both in the short- and long-term. If certain conditions are met (such as formal citation and sharing reprints) respondents agree they are willing to share their data. There are also significant differences and approaches in data management practices based on primary funding agency, subject discipline, age, work focus, and world region. Barriers to effective data sharing and preservation are deeply rooted in the practices and culture of the research process as well as the researchers themselves. New mandates for data management plans from NSF and other federal agencies and world-wide attention to the need to share and preserve data could lead to changes. Large scale programs, such as the NSF-sponsored DataNET (including projects like DataONE) will both bring attention and resources to the issue and make it easier for scientists to apply sound data management principles.

29. Trinidad, S.B., Fullerton, S.M., Bares, J.M., Jarvik, G.P., Larson, E.B., & Burke, W. (2010). Genomic research and wide data sharing: Views of prospective participants. *Genet Medicine, 12*(8), 486-495. doi:10.1097/GIM.0b013e3181e38f9e

    Abstract: This study was designed to explore the perceptions, beliefs, and attitudes of research participants and possible future participants regarding genome-wide association studies and repository-based research. Focus group sessions with (1) current research participants, (2) surrogate decision-makers, and (3) three age-defined cohorts (18-34 years, 35-50, >50). Participants expressed a variety of opinions about the acceptability of wide sharing of genetic and phenotypic information for research purposes through large, publicly accessible data repositories. Most believed that making de-identified study data available to the research community is a social good that should be pursued. Privacy and confidentiality concerns were common, although they would not necessarily preclude participation. Many participants voiced reservations about sharing data with for-profit organizations. Trust is central in participants' views regarding data sharing. Further research is needed to develop governance models that enact the values of stewardship.

30. Waaijers, L., & Van der Graaf, M. (2011). Quality of research data, an operational approach." *D-Lib Magazine, 17*(½). doi:10.1045/january2011-waaijers

    Abstract: The study investigated the operational aspects of the concept of quality for the various phases in the life cycle of research data: production, management, and use/re-use. Nine potential recommendations for quality improvement were derived from interviews and a study of the literature. The desirability and feasibility of these recommendations were tested by means of a national survey of university professors and senior lecturers, with a distinction being made in this regard between the three disciplinary domains applied by the European Science Foundation: Physical Sciences and Engineering, Social Sciences and Humanities, and Life Sciences.

31. Wolf, A., Simpson, M., Salo, D., Flee, D., Cheetham, J., & Barton, B. (2009). Summary report of the research data management study group. Retrieved August 1, 2013 from the WWW: http://minds.wisconsin.edu/handle/1793/34859

    Abstract: The Research Data Management Study Group (RDMSG) conducted focused interviews with representatives from a number of research communities, to assess current researcher data assets, needs, and funding situations. The interviews revealed a broad diversity in asset content and format, a large number of disparate needs, and an inadequate funding base for many researchers. The study group proposes a one-year pilot project to address the most common, most urgent subset of these issues.

Websites

1. ACRID (Advanced Climate Research Infrastructure for Data). Retrieved August 1, 2013 from the WWW: http://www.cru.uea.ac.uk/cru/projects/acrid/

   Description: Developed a linked-data approach to citing and publishing climate research data along with full provenance information, including the workflows and what software was used.

2. Alliance for Permanent Access. Retrieved August 1, 2013 from the WWW: http://www.alliancepermanentaccess.org/

3. CENS. Retrieved August 1, 2013 from the WWW: http://research.cens.ucla.edu/

4. CIRES. Retrieved August 1, 2013 from the WWW: http://cires.colorado.edu/

5. CMIP5. Retrieved August 1, 2013 from the WWW: http://cmip-pcmdi.llnl.gov/cmip5/

6. Creative Commons. http://creativecommons.org/ Creative Commons 0. Retrieved August 1, 2013 from the WWW: http://creativecommons.org/choose/zero/

7. CrossRef. Retrieved August 1, 2013 from the WWW: http://www.crossref.org/

8. Datacurate. Retrieved August 1, 2013 from the WWW: http://www.datacurate.com/

9. Data Curation Profiles Directory. Retrieved August 1, 2013 from the WWW: http://docs.lib.purdue.edu/dcp/

10. Data Interactive Publications. Retrieved August 1, 2013 from the WWW: https://sites.google.com/site/datainteractivepublications/

11. Data Seal of Approval (Ball & Duke, UKLON). Retrieved August 1, 2013 from the WWW: http://www.datasealofapproval.org/?q=node/66

12. Dataverse Network Citation Standard. Retrieved August 1, 2013 from the WWW: http://thedata.org/citation

13. DigCurV (2010). Retrieved August 1, 2013 from the WWW: Digital Curator Vocational Education Europe. http://www.digcur-education.org/

14. e-Bank. Retrieved August 1, 2013 from the WWW: http://www.ukoln.ac.uk/projects/ebank-uk/data-citation/

15. Elsevier. Retrieved August 1, 2013 from the WWW: www.elsevier.com

16. ESDS (Economic and Social Data Service). Retrieved August 1, 2013 from the WWW: http://www.esds.ac.uk/international/news/news.asp#21sep11

17. EPIC. Retrieved August 1, 2013 from the WWW: http://www.pidconsortium.eu/

18. EZID. Retrieved August 1, 2013 from the WWW: http://n2t.net/ezid

19. Federal Register. Retrieved August 1, 2013 from the WWW: https://www.federalregister.gov/articles/2011/11/04/2011-28621/request-for-information-public-access-to-digital-data-resulting-from-federally-funded-scientific

20. Guide to publication policies of the Nature Journals. Retrieved August 1, 2013 from the WWW: http://www.nature.com/authors/policies/availability.html.

    PDF here: Retrieved August 1, 2013 from the WWW: http://www.nature.com/authors/gta.pdf

21. ICPSR (Inter-university Consortium for Political and Social Research) Data Citations. Retrieved August 1, 2013 from the WWW: http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.js

22. International DOI Foundation. Retrieved August 1, 2013 from the WWW: http://www.doi.org/

23. MMI. Retrieved August 1, 2013 from the WWW: http://marinemetadata.org/

24. NERC. Retrieved August 1, 2013 from the WWW: http://www.nerc.ac.uk/research/sites/data/

25. OakLawProject. Retrieved August 1, 2013 from the WWW: http://www.oaklaw.qut.edu.au/

26. OECD (Organisation for Economic Co-operation and Development) (T. Green). Retrieved August 1, 2013 from the WWW: http://www.oecd.org/home/0,3675,en_2649_201185_1_1_1_1_1,00.html

27. Open Citations Corpus. Retrieved August 1, 2013 from the WWW: http://opencitations.net

    Blog here: Retrieved August 1, 2013 from the WWW: http://opencitations.wordpress.com

28. Openwetware. Retrieved August 1, 2013 from the WWW: http://openwetware.org/wiki/Main_Page

29. Plos. Retrieved August 1, 2013 from the WWW: http://www.plos.org/

30. The Scholarly Contributions and Roles Ontology (SCoRO). Retrieved August 1, 2013 from the WWW: http://purl.org/spar/scoro/

    Scholarly Contributions Report Form (SCORF) here: Retrieved August 1, 2013 from the WWW: http://purl.org/spar/scoro/scorf/ (Excel)

31. SPAR. Retrieved August 1, 2013 from the WWW: http://sempublishing.sourceforge.net/

32. SPQR. Retrieved August 1, 2013 from the WWW: http://spqr.cerch.kcl.ac.uk/ (Supporting Productive Queries for Research)

    Description: Trialed the use of linked data to express and integrate datasets related to classical antiquity, as a way of overcoming the challenges raised by the interpretive and uncertain nature of the material.

33. STFC. Retrieved August 1, 2013 from the WWW: http://www.stfc.ac.uk/e-Science/default.aspx

34. STM (International Association of Scientific, Technical & Medical Publishers). Retrieved August 1, 2013 from the WWW: http://www.stm-assoc.org/about-the-association/

35. Thompson Reuters. Retrieved August 1, 2013 from the WWW: http://researchanalytics.thomsonreuters.com/solutions/researcherid/

36. USGS data management. Retrieved August 1, 2013 from the WWW: http://www.usgs.gov/datamanagement/describe/citation.php

37. University of Oxford Research Data Management. Retrieved August 1, 2013 from the WWW: http://www.admin.ox.ac.uk/rdm/

38. W3C Open Annotation. Retrieved August 1, 2013 from the WWW: http://www.openannotation.org/spec/core/

39. Webtracks. Retrieved August 1, 2013 from the WWW: http://www.stfc.ac.uk/e-Science/projects/medium-term/metadata/webtracks/22422.aspx

    Description: Extended previous work by the CLADDIER and StoreLink projects in order to produce a secure method for communicating semantic links between data repositories, publication repositories, open science notebooks and publishers.

40. XYZ Project. Retrieved August 1, 2013 from the WWW: http://projectxyz.wordpress.com/