CODATA Task Group on Digital Data Citation - Best Practices: Research & Analysis Results

As part of this year's activities of the CODATA Digital Data Citation Task Group, we conducted an inventory of existing literature as well as data citations and attribution activities. The idea behind this effort was to collect sources of information related to how data repositories cite and provide attributions to their data sets. This document is the result of the collection of bibliographic sources, subsequent research and corresponding analysis.

The collection was created by members of the group and consultants capturing information sources that are directly or peripherally focused on digital data citation practices and attribution. These contributions were made via email or the Zotero tool. Additional sources were discovered through online searches.

We found 384 resources in 15 different formats that covered 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 5 years with the occasional older seminal item included because they provided additional context and background to writers of the white paper on the best practices and standards in attribution and citation of scientific data. Each source contains links and notes or abstracts where applicable/possible. Research papers comprise the bulk of the bibliography and we classified those into research papers, government & committee reports and surveys & 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	22	5.73%
Books	10	2.60%
Citation Guides	33	8.59%
Citation Software & Repositories	44	11.46%
Conferences, Workshops, Symposia, Meetings	13	3.39%
Journal issues devoted to data	3	0.78%
Op-eds, Newsletters, Press Releases, Memorandums	13	3.39%
Organizations, Committees	24	6.25%
Papers	111	28.91%
Papers: Government, Committee	20	5.21%
Papers: Surveys, Studies	30	7.81%
Posters, Charts	6	1.56%
Presentations PPTs, Videos	16	4.17%
Standards	9	2.34%
Websites	30	7.81%
Grand Total	**384**	**100.00%**

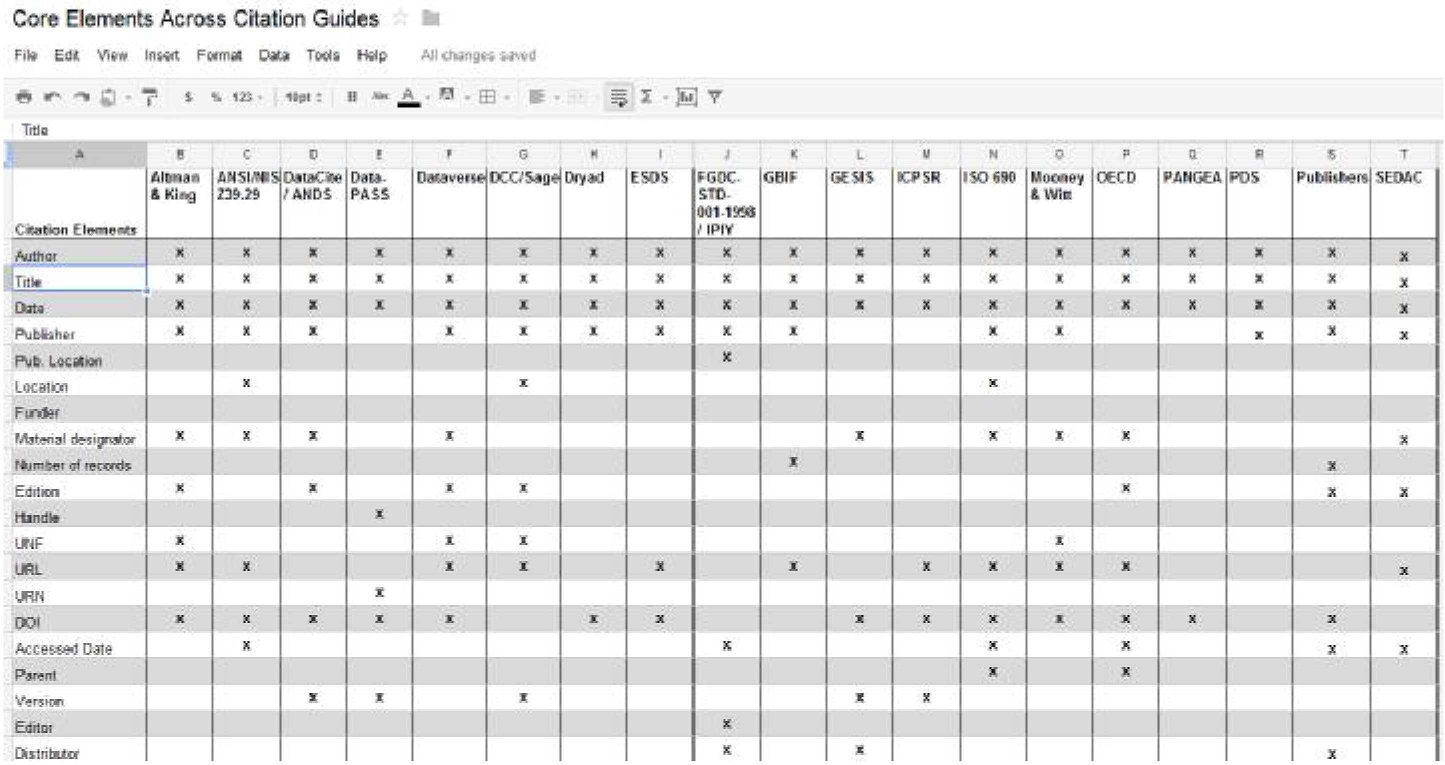
The topics covered the most by the literature include:

- 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 did not find a great number of policy standards applicable to digital data citations, neither did we find a consensus practice (or practices) for data attribution. We found scattered best practices and varied among disciplines, when available. The data citation practices that we yielded as part of the research trend toward traditional print data citation methods and not 21st Century scientific digital data.

From our review of the literature we found a number of citation guidelines, some of which we consensus practices or best practices for an organization. We examined core elements across citation guidelines in our bibliography and created a chart representing our findings – see table below (also see the original document following the Google Docs URL https://docs.google.com/spreadsheet/ccc?key=0ArU4DBwxYrfAdDJMWnhGN1hJTGg2SmRrOFRyTHRXZHc.) We found a wide range of practices regarding required elements in citation practices.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Citation Elements	Altman & King	ANSI/NIS Z39.29	DataCite / ANDS	Data-PASS	Dataverse	DCC/Sage	Dryad	ESDS	FGDC-STD-001-1998 / IPIY	GBIF	GESIS	ICPSR	ISO 690	Mooney & Wiz	OECD	PANGEA	PDS	Publishers	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	

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 regards to assisting researchers with data management.

The resources of our collection are going to provide documentation support for writing the best practices whitepaper, we have incorporated the outline as Appendix A. Under each of the 6 main topics of the outline, we inserted 100+ resources that we considered the most relevant to each topic. Appendix B is the full bibliography. Our goal is to further populate the collection and provide additional context that speak to the main topics of the outline, with a primary focus on papers, reports, and surveys from the bibliography. This process will continue until the paper is published.

A suggested next step to continue the bibliographic documentation research is to assess the feasibility of best practice needs focusing on sharing practices, differences and similarities among different scientific disciplines – including what policy and incentives are applied and could be shared in digital data citations.

Appendix A

Overview and Current Practices for Data Citation

(104 suggested sources added)

1. **Importance of data citation**
   a. Increased importance of data management, sharing, replication
      i. Data's role in the research life cycle
         1. Data sets as first class research products (introduce, see also later)

Callaghan, C., Donegan, S, Pepler, S. Thorley, M., Cunningham, N., Kirsch, P. 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). Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/208 [see papers #34]

Heery, R. (2009). "Digital Repositories Roadmap Review: towards a vision for research and learning in 2013." Retrieved from http://www.jisc.ac.uk/media/documents/themes/infoenvironment/reproadmapreviewfinal.doc [see papers #62]

Waaijers, L. and Van der Graaf, M. (2011). "Quality of Research Data, an Operational Approach." *D-Lib Magazine January/February 2011 Volume 17, Number ½.* Retrieved from http://www.dlib.org/dlib/january11/waaijers/01waaijers.html [see surveys and studies #29]

      ii. Institutional recognition of formal need for data management--
         1. Definitions: internal management, short-term dissemination/sharing, long term access
         2. Long-lived data collection NSF report; Blue Ribbon Task Force Report on Preservation; Data management plan requirements introduced by funders; Increasing publisher focus on management of "supplementary materials"

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 from http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf [see reports #2]

Long-lived data collection NSF report http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf

      iii. Disciplinary movements towards data sharing --
      positive benefits (open data movement, nature/science editorials, democratization of data access, increase in impact, reuse); desire to avoid negatives associated with data inaccessibility (replication, increase in retractions, research integrity)

Cook, R. (2008). "Citations to published data sets." *FLUXNET Newsletter*. http://daac.ornl.gov/ornl_daac_citations_200812.pdf [see op-eds, newsletters, press releases, memorandums #3]

Costello, M. J. 2009. Motivating online publication of data. *Bioscience 59 (5): 418-427*. Retrieved from http://www.jstor.org/discover/10.1525/bio.2009.59.5.9?uid=3739912&uid=2&uid=4&uid=3739256&sid=55925848753 [see papers #41]

Nelson, B. (2009). "Data sharing: Empty archives." *Nature 461:160-163.* Retrieved from http://www.nature.com/news/2009/090909/full/461160a.html [see papers #81]

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*. Retrieved from http://www.springerlink.com/index/10.1007/BF02628694 [see papers #97]

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 [see papers #101]

Vision, T.J. (2010). "Open data and the social contract of scientific publishing." *American Institute of Biological Sciences, 60(5), 330-331*. Retrieved fromhttp://caliber.ucpress.net/doi/abs/10.1525/bio.2010.60.5.2 [see papers #106]

b. Increasing complexity of data

. data deluge
-- production of research data growing geometrically

Bohn, R., Short, J. (2009). "How Much Information? 2009 Global Information Industry Center Report on American Consumers." Retrieved from http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf [see reports #3]

Borgman, C. (2011). "The conundrum of sharing research data." *Journal of the American Society for Information Science and Technology, pp. 1-40, 2011*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155 [see papers #25]

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 from http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf [see papers #53]

Hey, T., Trefethen, A. (2003). "The data deluge: An e-science perspective." *From "Grid Computing – making the global infrastructure a reality", Wiley*. Retrieved from http://eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf [see papers #65]

i. shifting evidence base: distributed/social production of knowledge
-- data production, collection, and management increasingly moved from large/central production to distributed groups, individuals

ii. shifting evidence base: new forms of data
-- within fields, data formats and sources are expanding, e.g. to crowd sourced data entry, mobile phone data collection, social networks, and other non-traditional forms of research evidence

c. Role of data citation as a key part of infrastructure required for data management, sharing, replication, research integrity

Borgman, C. (2007). "Scholarship in the Digital Age: Information, Infrastructure, and the Internet." *The MIT Press.* Retrieved from http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=11333 [see books #2]

Berman, F. (2010). "We Need a Research Data Census." *Communications of the ACM Vol. 53 No. 12, Pages 39-41*. http://cacm.acm.org/magazines/2010/12/102121-we-need-a-research-data-census/fulltext [see op-eds, newsletters, press releases, memorandums #1]

High Level Expert Group on Scientific Data (2010). “Riding the Wave: How Europe Can Gain from the Rising Tide of Scientific Data. European Commission.” Retrieved from http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf [see reports #8]

National Science Foundation (2011). “Advisory Committee for Cyberinfrastructure, and Task Force on Data and Visualization. Final Report.” *Arlington, VA: National Science Foundation*. Retrieved from http://www.nsf.gov/od/oci/taskforces/TaskForceReport_Data.pdf [see reports #15]

Fitzgerald, A. Pappalardo, K. (2007). “Building the infrastructure for data access and reuse in collaborative research.” Retrieved from http://eprints.qut.edu.au/8865/1/8865.pdf [see papers #48]

Johnston, L. (2010). “User-needs assessment of the research cyberinfrastructure for the 21st century.*”* *Perdue University*. Retrieved from http://docs.lib.purdue.edu/iatul2010/conf/day1/5/ [see surveys and studies #18]

National Science Foundation (2011). “Advisory Committee for Cyberinfrastructure, and Task Force on Data and Visualization. Final Report.” *Arlington, VA: National Science Foundation*. Retrieved from http://www.nsf.gov/od/oci/taskforces/TaskForceReport_Data.pdf [see reports #15]

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 Retrieved from http://www.agu.org/pubs/crossref/2010/2010EO340001.shtml [see papers #83]

Paton, N.W. (2008). “Managing and sharing experimental data: standards, tools and pitfalls.” *Biochemical Society Transactions 36 (1), 33-36*. Retrieved from http://www.mendeley.com/research/managing-and-sharing-experimental-data-standards-tools-and-pitfalls/ [see papers #86]

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, 2005, Volume 3652/2005, 128-138.* Retrieved from http://www.springerlink.com/content/2u3eng7kvt58t7v9/ [see papers #93]

A key part of supporting a range of uses:

1. attribution --
   legal attribution and scientific credit (which are not the same)
2. persistence --
   persistence of reference; identity of curators responsible for data set (need to associate role with individual who currently occupies that role)

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 [see reports #5]

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.* :1-22. Retrieved from http://dx.doi.org/10.1007/s12145-011-0083-6 [see papers #47]

Hakala, J. (2010). “Persistent identifiers – an overview.” *The KIM Technology Watch Report*http://metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/ [see papers #59]

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.* Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/18445641 [see papers #82]

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 Issue 1, Volume 3* . Retrieved from http://www.ijdc.net/index.php/ijdc/article/viewFile/67/46 [see papers #109]

Wynholds, L. (2011). “Linking to scientific data: Identity problems of unruly and poorly bounded digital objects.” *International Journal of Digital Curati*on 6(1).Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/174 [see papers #111]

3. access --
   short & long term; machine & human
4. discovery --
   locate instances; discover derivative/parent/citing works
5. provenance --
   associate scientific claim and specific evidence; verify fixity of evidence

Brase, J., Farquhar, A., Gastl, A., Gruttemeier, H., Heijne, M., Heller, A., Hitson, B., Johnson, L., McMahon, B., Piguet, A., Rombouts, J., Sandfaer, M., & Sens, I. (2009). “Numeric Data: Citation Techniques and Integration with Text.” Retrieved from http://www.icsti.org/IMG/pdf/Numeric_Data_FINAL_report.pdf [see papers #29]

Cheney, J., Chiticariu, L., Tan,W.-T. (2009). “Provenance in databases: Why, where and how.” *Foundations and Trends® in Databases: Vol. 1: No 4, pp 379-474.* Retrieved from http://www.nowpublishers.com/product.aspx?product=DBS&doi=1900000006 [see papers #37]

Freire, J., Koop, D., Santos, E., Silva, C. (2008). “Provenance for Computational Tasks: A Survey.” *Computing Science and Engineering, Vol 10, No 3, pp 11-21, 2008.* Retrieved from http://www.computer.org/portal/web/csdl/doi/10.1109/MCSE.2008.79 [see papers #49]

Moreau, L. (2010). “The Foundations for Provenance on the Web.” *Foundations and Trends® in Web Science: Vol. 2: No 2-3, pp 99-241*. Retrieved from http://eprints.soton.ac.uk/271691/1/survey.pdf [see papers #80]

Simmhan, Y., Plale, B., Gannon, D. (2005). “A survey of data provenance in e-science.” *ACM SIGMOD Vol 34, No 3, 2005*. Retrieved from http://pti.iu.edu/sites/default/files/simmhanSIGMODrecord05.pdf [see surveys and studies #24]

Tilmes, C., Yesha, Y., Halem, M. (2011). “Distinguishing Provenance Equivalence of Earth Science Data.” *Procedia Computer Science Volume 4, 2011, Pages 548–557*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1877050911001153 [see papers #104]

W3C. Incubator report (2010). Retrieved from http://www.w3.org/2005/Incubator/prov/XGR-prov-20101214/. [see papers #107]

i. Definitions -- What are citations, citation components, extended citations? Analogy to literature citations is not a very complete match. Need to distinguish practices and functions supported by citations from citation format per se.

in-text reference; persistent identifier; bibliographic reference; extended metadata bound to citation in common catalog like crossref; external catalog information

ii. Must be effectively integrated into scholarly communication ecosystem:
research design, data collection, analysis; research funding; data archiving & dissemination; scholarly publication; tenure and promotion

iii. Effects a variety of stakeholder:
researchers as data collectors, authors of articles, users of secondary data; journal editors; journal publishers; research institutions as data managers; funders; librarians; tenure and promotion committees; data publishers; data repositories, centers, archives

Current use of Data Citations

. *Exemplary Data Repositories/Publishers:*
A number of repositories/data publishers have developed good, consistent practice, examples illustrate these, though specifics vary, and list is not comprehensive:

BMC BL Data repositories. Lists 155 domain-specific and general data repositories. Includes name, website, subject area, funding model, restrictions, license agreement, county, identifiers, abbreviation, notes, representatives, and standardshttps://docs.google.com/spreadsheet/ccc?authkey=COmDvOUB&key=0Aok0Od_Hhd1XdEdiRXVCbDlFWk8wNW5FYlBBTndyaVE&hl=en_US&authkey=COmDvOUB#gid=0 [see posters and charts #1]

. ICPSR http://www.icpsr.umich.edu/icpsrweb/ICPSR/
  i. Pangea http://www.pangaea.de/
  ii. DataVerse http://thedata.org
  iii. Dryad http://datadryad.org/
    a. Incomplete practices and gaps

. inconsistent use of data citation by authors
  i. inconsistent treatment of data citation by editors
  ii. inconsistent use by catalogs

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 from http://www.dcc.ac.uk/webfm_send/303 [see posters and charts #2]

Newton, M. Mooney, H, Witt, M. “A Description of Data Citation Instructions in Style Guides.” Retrieved from http://docs.lib.purdue.edu/lib_research/121/ [see posters and charts #3]

Piwowar,H. Chapman, W. (2007) “Examining the uses of shared data,” Poster.Retrieved fromhttp://precedings.nature.com/documents/425/version/2/html [see posters and charts #6]

Emerging formal standardization proposals and best practices development

General standards/practices development: DataCite http://datacite.org/;

OECD http://www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html;

. Data-PASS/Dataverse; http://www.data-pass.org/ http://thedata.org/
DCC http://www.dcc.ac.uk/

a. Exemplary disciplinary efforts:
SageCite, http://www.ukoln.ac.uk/projects/sagecite/
GBIF http://www.gbif.org/
Federation of Earth Science Information Partners http://www.esipfed.org/

Data Paper efforts

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 from http://archivists.metapress.com/content/EU7252LHNRP7H188 [see papers #6]

Callaghan, C., Donegan, S, Pepler, S. Thorley, M.,Cunningham, N., Kirsch, P. 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). Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/208 [see papers #34]

Lane, M. (2008). "Data citation in the electronic environment." A white paper commissioned by GBIF. Retrieved from http://www.danbif.dk/Documents/gbif-documents/DataCitation-Lane2008.pdf [see papers #72]

Emerging Principles for Data Citation

. General Scientific Principles

. The published article is (only) a summary of the research

i. The published article provides context for a data set

ii. Science requires reproducibility

iii. Disciplines require a shared evidence base

a. Core Requirements

. Data citations should be "first class objects" *for publication* -- appear in references; be as easy to reference as other works

i. All evidence (including data) necessary to assess conclusions in scholarly work should be cited

ii. Citations should persist, and enable access to fixed/intended version of data, as long as the citing work exists

iii. Citation should *support* attribution of credit to all contributors (possibly indirectly, through citation ecosystem, medata, indices)

Callaghan, C., Donegan, S, Pepler, S. Thorley, M., Cunningham, N., Kirsch, P. 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). Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/208 [see papers #34]

b. Design principles

. separate scientific principles, use cases technical requirements

i. Distinguish syntax from presentation

ii. Design for ecosystem and lifecycle

iii. Incremental value for incremental effort -- simple & weak

iv. Scalable

v. Open

Rodriguez, M., Bollen, J., Sompel, H. (2007). "A Practical Ontology for the Large-Scale Modeling of Scholarly Artifacts and their Usage." *In Proceedings of the Joint Conference on Digital Libraries, Vancouver, June 2007.* Retrieved from http://public.lanl.gov/herbertv/papers/Papers/2007/JCDLrodriguez.pdf [see papers #91]

c. Operational requirements for fields in semantic citation
(not necessarily in particular presentation of citation)

Bernstein, H. J., Folk, M. J., Benger, W., Dougherty, M. T., Eliceiri, K. W. and Schnetter, E. (2011). “Communicating Scientific Data from the Present to the Future. Dowling College position paper.” Temporary URL: http://www.columbia.edu/~rb2568/rdlm/Bernstein_Dowling_RDLM2011.pdf [see papers #16]

National Science Foundation (2011). “Digital research data sharing and management.” Retrieved fromhttp://www.nsf.gov/nsb/publications/2011/nsb1124.pdf [see reports #17]

. *Include a persistent identifier*

i. When citation is presented in a electronic context (like a web browser) provide a actionable reference (e.g. a link)

ii. Include an author (or corporate author) -- need not include all contributors in citation itself

iii. Include a title: even if generic

iv. Include a version or quasi-version:
in preference order: formal version, date modified, date accessed

v. Field-specific practices
- need to be permitted by flexible data citation practice, but requirement will vary

vi. granularity of citation -- reference to appropriate piece of work

vii. types of direct & indirect attribution -- when should scholarly attribution appear in in-text reference, extended metadata, accompany data -paper, etc. (e.g. Galaxy Zoo has 200K “contributors”)

viii. syntax and presentation: wide variety of citation styles, formats, both on-line and in print

ix. types of versioning information --
usually provenance requires reference to specific version of evidence -- but wide variety of versioning approaches including embedding in identifier; extended reference; data of last change; formal version numbers

x. cite to non-versioned/dynamic work --
on occasion one intends to cite “most current version of” or “general data collection” -- e.g. when data is cited as part of a review article/for teaching/ and not as evidentiary support

xi. semantic validation of data and file format-indepedent citation --
semantics of data are logically separate from format; in some cases semantic fingerprints are available and data can be cited independent of file format; in others they are not easily separable and format of data must be indicated in version or extended citation information

Altman, M. (2008). “A Fingerprint Method for Verification of Scientific Data.” *A Fingerprint Method for Verification of Scientific Data. : Springer-Verlag*. Retrieved from http://thedata.org/publications/fingerprint-method-verification-scientific-data [see papers #7]

Altman, M., & King, G. (2007). “A proposed standard for the scholarly citation of quantitative data.” *D-Lib Magazine, 13*(3/4). Retrieved from http://gking.harvard.edu/files/abs/cite-abs.shtml [see papers #9]

Bollen, J., Sompel, H. (2006). “An Architecture for the Aggregation and Analysis of Scholarly Usage Data.” *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries.* Retrieved from http://arxiv.org/abs/cs/0605113 [see papers #17]

Buneman, P. Silvello, G (2010). “A Rule-Based Citation System for Structured and Evolving Datasets.” *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*. Retrieved from http://sites.computer.org/debull/A10sept/buneman.pdf [see papers #33]

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 Issue 2, Volume 6 | 2011.* Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/181/265 [see papers #73]

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 January/February 2011 Volume 17, Number ½.* Retrieved from http://www.dlib.org/dlib/january11/michener/01michener.html [see papers #79]

d. Technical/operational requirements

Tools and Infrastructure

. Current situation

a. Overview of needs

. formal standards needed

i. best practices, documentation, curricula needed

Autodesk Geospatial (2007). “Best Practice for Managing Geospatial Data.” Retrieved from http://www.gisperfect.com/res/AutocadMAP/best_practices.pdf [see papers #14]

Brown, D., Welch, G., Cullingworth, C. (2005). “Archiving, management and preservation of Geospatial data.” Retrieved fromhttp://www.geoconnections.org/publications/policyDocs/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf [see papers #31]

Buneman, P. Silvello, G (2010). “A Rule-Based Citation System for Structured and Evolving Datasets.” *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*. Retrieved from http://sites.computer.org/debull/A10sept/buneman.pdf [see papers #33]

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 (Suppl14), S2.* Retrieved from http://www.biomedcentral.com/1471-2105/10/S14/S2 [see papers #36]

CIESIN Columbia University (2005). “Data model for Manafing and preserving Geospatial Electronic Records.” Retrieved from http://www.ciesin.columbia.edu/ger/DataModelV1_20050620.pdf [see papers #38]

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 from www.daac.ornl.gov/DAAC/PI/bestprac.html#prac2

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 from http://daac.ornl.gov/PI/BestPractices-2010.pdf [see papers #66]

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 from http://escholarship.org/uc/item/9jw4964t#page-2 [see papers #71]

“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 from http://www.geoinfogeochem.org/sites/geoinfogeochem.org/files/Policy_GeochemDataPubl_v1.1_0.pdf [see papers #105]

ii. technical infrastructure needed:
cataloging and indexing (e.g. data citations in crossref); citation management tools; extensions to workflow systems and repository tools; extension to manuscript management systems

Cultural Challenges and Opportunities

. *Challenges vary by discipline*

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 from http://docs.lib.purdue.edu/iatul2010/conf/day1/1/ [see papers #11]

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 (2002) Volume: 6, Issue: 2, Pages: 241.* Retrieved from http://www.mendeley.com/research/datasharing-datawithholding-genetics-life-sciences-results-national-survey-technology-transfer-officers-1/ [see papers #35]

Lowry, R., Urban, E., & Pissierssens, P. (2009). “A New Approach to Data Publication in ocean sciences.”*Eos, Vol. 90, No.* 50.http://www.agu.org/pubs/crossref/2009/2009EO500004.shtml [see op-eds, newsletters, press releases, memorandums #9]

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* Fall 2011 DOI:10.5062/F4CC0XMJ. Retrieved from http://www.istl.org/11-fall/article1.html [see papers #76]

Parsons, M., Bruin, T., Tomlinson, S., Campbell, H., Godoy, O., LeClert, J.,et al.(2009). “The State of Polar Data—the IPY Experience.” Retrieved from http://ipydis.org/documents/State_of_Polar_Data20100514_distribute.pdf [see papers #76]

Research information network. (2011). “Physical Sciences Case studies: information use and discovery.” Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/physical-sciences-case-studies-use-and-discovery- [see papers #84]

Research information network. (2011). “Reinventing research? Information practices in the humanities.” Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities [see reports #21]

Thessen, A., Patterson, D. (2011). “Data issues in the life sciences.” *White paper*. Retrieved from http://dataconservancy.org/sites/default/files/Data%20Issues%20in%20the%20Life%20Sciences%20White%20Paper.pdf [see papers #103]

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 Med. 2010 Aug;12(8):486-95*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20535021 [see surveys and studies #28]

Waaijers, L. and Van der Graaf, M. (2011). “Quality of Research Data, an Operational Approach.” *D-Lib Magazine January/February 2011 Volume 17, Number ½.* Retrieved from http://www.dlib.org/dlib/january11/waaijers/01waaijers.html [see surveys and studies #29]

a. Commitments by stakeholder groups

Pinowar, H. Day, R. Fridsma, D. (2007) “Sharing detailed research data is associated with increased citation rate.” http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000308 [see surveys and studies #15]

b. Changing perceptions and environments
c.

Open research questions

. Scientific questions:
identifiying integral vs. ancillary data; minimum information needed for reproducibility in particular fields; selection of data for long-term access/storage -- reuse potential; canonicalization of common data objects -- semantic definition of data in particular fields

a. Technical questions: (see tools and infrastructure needs)
b. Institutional (legal/financial, organizational) questions & roles:

Borgman, C. (2007). “Scholarship in the Digital Age: Information, Infrastructure, and the Internet.” *The MIT Press.* Retrieved from http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=11333 [see books #2]

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/ [see papers #90]

. role of publisher --
robust connection of article and data w/out requiring publisher to archive all data as supplementary materials; integration of data publishing and journal publishing workflow; indexing data and articles together; connection of author id’s and data

Aalbersberg, I. and Kahler, O. (2011). “Supporting Science through the Interoperability of Data and Articles.” *D-Lib Magazine January/February 2011 Volume 17, Number ½.* http://www.dlib.org/dlib/january11/aalbersberg/01aalbersberg.html#3 [see papers #1]

Green, T. (2009). “We need publishing standards for datasets and data tables.” *OECD Publishing White Paper, OECD Publishing*. Retrieved from http://dx.doi.org/10.1787/603233448430 [see papers #57]

Maunsell, J. (2010). “Announcement regarding supplemental material.” *The Journal of Neuroscience 11 August 2010, 30(32): 10599-10600.* Retrieved from http://www.jneurosci.org/content/30/32/10599.full [see op-eds, newsletters,press releases, memorandums#10]

National Information Standards Organization (NISO), National Federation of Advanced Information Services (NFAIS) (2010). “Roundtable on Best Practices for Supplemental Journal Article Materials.” Retrieved from http://iassist-sigdc.googlegroups.com/attach/7186703f23266e75/RP-15-201x+Suppl_BWG_draft_for_comments.pdf?view=1&part=2 [see reports #13]

Piwowar, H., Chapman, W. (2008). A review of the journal policies for sharing research data. In *ELPUB*. Retrieved from http://ocs.library.utoronto.ca//index.php/Elpub/2008/paper/view/684 [see surveys and studies #14]

PR Newswire (2010). “Elsevier and PANGAEA Take Next Step in Connecting Research Articles to Data.” *United Business Media*. Retrieved from http://www.prnewswire.com/news-releases/elsevier-and-pangaea-take-next-step-in-connecting-research-articles-to-data-99533624.html [see op-eds, newsletters, press releases, memorandums #10]

role of editors --
best editorial practice for replication and citation; workflow support -- role of copyeditor vs. author in data citation

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, 0 DOI: 10.1214/10-STS320. Retrieved from http://arxiv.org/pdf/1011.0810v1.pdf [see papers #95]

Smit, E. (2011). "Avoiding a Digital Dark Age for data: why publishers should care about digital preservation." *Learned publishing* 24(1), 35-49. Retrieved from http://www.ingentaconnect.com/content/alpsp/lp/2011/00000024/00000001/art00007 [see papers #98]

i. role of tenure and review committee --
evaluating impact of published data

ii. funders --
what should be recommended/required in data management plan wrt to citation; what should be recommended/required in publication of articles related to research; how should compliance with requirements be evaluated by funders

Jones, S. (2012). "Developments in research funder data policy." *International Journal of Digital Curation* 7(1). Retrieved from http://ijdc.net/index.php/ijdc/article/view/209/278 [see papers #68]

National Science Foundation (2011). "Digital research data sharing and management." Retrieved from http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf [see reports #17]

National Science Foundation (2011). "Division of Ocean Sciences Sample and Data Policy." Retrieved from http://www.nsf.gov/pubs/2011/nsf11060/nsf11060.pdf [see reports #18]

Organization for Co-operation and Development (2007). "OECD Principles and Guidelines for Access to Research Data from Public Funding." Retrieved from http://www.oecd.org/dataoecd/9/61/38500813.pdf [see reports #20]

iii. librarians --

what should librarians preserve for the longer term; how should data sets be captured in library catalogues; how could best citation practice be reinforced through literacy training;

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 from http://www.box.net/shared/d3cf8u0gtyml2nqq3u2f [see papers #5]

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 from http://docs.lib.purdue.edu/iatul2010/conf/day1/1/ [see papers #11]

Brase, J. (2004). "Using Digital Library Techniques- Registration of Scientific Primary Data." *Research and Advanced Technology for Digital Libraries 8th European Conference, ECDL 2004, Bath, UK, September 12-17, 2004. Proceedings.* Retrieved from http://www.springerlink.com/content/1pglbmjv95tqby9e/ [see papers #30]

vi. authors/researchers --

in their various roles as users of secondary data, producers of data

Cragin, M. H., Palmer, C. L., Carlson, J.R., and Witt, M. (2010). "Data sharing, small science and institutional repositories." *Phil. Trans. R. Soc. A 13 September 2010 vol. 368 no. 1926 4023-4038.* Retrieved from http://rsta.royalsocietypublishing.org/content/368/1926/4023 [see papers #42]

Piwowar, H., Chapman, W. (2010). "Public sharing of research datasets: A pilot study of associations." *148-156. In Journal of Informetrics 4 (2).* Retrieved from http://www.sois.uwm.edu/MetricsPreCon/documentation/Piwowar_Chapman_Sharing.pdf [see surveys and studies #13]

Research information network., (2011). "Information handling in collaborative research: an exploration of five case studies." Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/collaborative-research-case-studies [see surveys and studies #19]

Research Information Network (2008). "To share or not to share." Retrieved from http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs [see reports #21]

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*. Retrieved from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0021101 [see surveys and studies #27]

d. Scientific culture norms and practices:
- what are the range of and best practice examples of field specific examples in 4 d, above: granularity, syntax and presentation, versioning, dynamic works, semantic validation/format independent citation

Helliwell, J. R. and McMahon, B. (2010). "The record of experimental science: Archiving data with literature." Retrieved from http://iospress.metapress.com/content/f0765625774j4051/fulltext.pdf [see papers #63]

e. Bibliometric/impact:
- measurement of data impact via citation vs. download and other measures of use; effect of data citation on overall impact ; are fields connected through data that are not connected through publication?

Bollen, J., Rodriguez, M., Sompel, H. (2006). "Journal Status." *Scientometrics, volume 69, number 3, pp. 669-687, 2006.* Retrieved from http://arxiv.org/abs/cs.DL/0601030 [see papers #21]

Bollen, J., Sompel, H., HagBerg, A., Chute, R. (2009). "A principal component analysis of 39 scientific impact measures." *Cornell University Library*. Retrieved from http://arxiv.org/abs/0902.2183 [see papers #20]

Bollen, J., Sompel, H., Smith, J., Luce, R. (2005). "Toward alternative metrics of journal impact: a comparison of download and citation data.*" Information Processing & Management Volume 41 Issue 6 Pagination 1419-1440.* Retrieved from http://arxiv.org/abs/cs.DL/0503007 [see papers #22]

Bollen, J., Sompel, H., Rodriguez, M. (2008). "Towards Usage-based Impact Metrics: - First Results from the MESUR Project." *Proceedings of the Joint Conference on Digital Libraries 2008*. Retrieved from http://arxiv.org/abs/0804.3791 [see papers #23]

Bollen, J., 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 Volume 59 Issue 1, January 2008*. Retrieved from http://arxiv.org/pdf/cs.DL/0610154.pdf [see papers #24]

Pinowar, H. Day, R. Fridsma, D. (2007) “Sharing detailed research data is associated with increased citation rate.” http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000308 [see surveys and studies #15]

Appendix B

CODATA Bibliography

Blogs, Wikis, Web Groups

1. Altman, M. Blog (2012). “Micah Altman’s Blog.” Retrieved from http://drmaltman.wordpress.com/

2. Bibliographic Ontology Specification Group. Retrieved from http://groups.google.com/group/bibliographic-ontology-specification-group/about?hl=en
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. Callaghan, S. Blog (2012). “Citing Bytes.” Retrieved from http://citingbytes.blogspot.com/2011/12/idcc-2011-notes-from-day-1-plenary.html

4. DataCite Blog (2011). “Tracking Data Citation entry” Retrieved from http://datacite.wordpress.com/2011/01/15/tracking-data-citation/

5. DataCite Users Google Group. Retrieved from https://groups.google.com/forum/#!forum/datacite-users

6. Digital Preservation Matters Blog. Retrieved from http://preservationmatters.blogspot.com/2011/10/cite-datasets-and-link-to-publications.html

7. Dryad Wiki. “Data Citation Guidelines.” Retrieved from https://www.datadryad.org/wiki/Citing_Data

8. Earth Science Information Partner Federation. Wiki. “Data Stewardship/Citations” Retrieved from http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations/provider_guidelines

9. ESIP Federation (2012). “Interagency Data Stewardship/Citations/provider guidelines.” Retrieved from http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations/provider_guidelines

10. Gipp, B. Blog (2010). “JabRef + automatic metadata extraction from PDF files.” Retrieved from http://gipp.com/jabref-automatic-metadata-extraction-from-pdf-files-like-mendeley-2

11. Global Warming Policy Foundation: Best of Blogs (2011). “Joe Pickrell: Why Publish Science In Peer-Reviewed Journals*?” Genomes Unzipped, 13 July 2011*. Retrieved from http://thegwpf.org/best-of-blogs/3440-joe-pickrell-why-publish-science-in-peer-reviewed-journals.html

12. IASIST SIGDC (Special Interest Group on Data Citation). Google Group. Retrieved from http://groups.google.com/group/iassist-sigdc/browse_thread/thread/abc7c7b28e0df580
Promotes awareness of data-related research and scholarship through data citation. Includes style guides from Mooney and Witt’s poster session.

13. IDMB Blog. Retrieved from http://www.southamptondata.org/idmb-blog.html

14. IPAW Wiki. Retrieved from http://tw.rpi.edu/portal/Main_Page

15. Knowledge Blog. Retrieved from http://knowledgeblog.org/

16. OJIMS. Retrieved from http://proj.badc.rl.ac.uk/ojims

17. Piwowar, H. Blog (2011). “Resources on Data Citation Principles.” Research Remix blog posting. Retrieved from http://researchremix.wordpress.com/2011/05/17/resources-on-data-citation-principles

18. Roure, D. (2010). “Replacing the Paper: The Twelve Rs of the e-Research Record.” *R&D Information Services.* Retrieved from http://blogs.nature.com/eresearch/2010/11/27/replacing-the-paper-the-twelve-rs-of-the-e-research-record
Provides a 6-point definition of the properties of sharable Research Objects.

19. SageCite. Blog. Retrieved from http://blogs.ukoln.ac.uk/sagecite/
Produced a demonstrator citation service for network models, workflows and associated data in the Sage Commons, using a linked data approach.

20. Saller, C. (2011). “‘Citation Obsession’? Dream On.” *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/linguafranca/2011/11/03/citation-obsession-dream-on/

21. TWR: Standards in metadata. Retrieved from http://metadaten-twr.org

22. W3C Provenance Working Group Standardization Activity. Retrieved from 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: John Wiley & Sons*. Retrieved from http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471236330.html?0471236330
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 timation; integrated examples of statistical computing (using software packages such as the SAS, Gauss, Splus, R, Stata, 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. Borgman, C. (2007). “Scholarship in the Digital Age: Information, Infrastructure, and the Internet.” *The MIT Press.* Retrieved from http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=11333
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.

3. 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 from http://www.nap.edu/catalog.php?record_id=12615

4. Fetterer, F., H. Eicken (Ed.) (2009). "Data Management Best Practices for Sea Ice Observation**."** *Field Techniques for Sea-Ice Research, University of Alaska Press, ISBN 978-1-6022230-59-0.* Retrieved from http://nsidc.org/about/bios/fetterer.html
   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.

5. Geoscience Information Society, European Association of Science Editors (1999). "Science Editing and Information Management: Proceedings of the Second International Aese/ Cbe /Ease Joint Meeting." *Geoscience Information Society*. Out of Print; limited availability. Retrieved from http://www.amazon.com/dp/0934485305

6. Heath, T., Bizer, C. (2011). "Linked Data: Evolving the Web into a Global Data Space." Retrieved from http://linkeddatabook.com/book
   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.

7. Kowalczyk, S., Shankar, K. "Data sharing in the sciences." *Ch. 6, Annual review of information science and technology*. Retrieved from http://kalpanashankar.files.wordpress.com/2010/06/arist_data_sharing.pdf

8. Murphy, C., (1982). "Micrometeorological Data for the Energy Balance and the Exchange of Carbon Dioxide between a Forest and the Atmosphere." Print on demand. Retrieved from http://www.ntis.gov/search/product.aspx?ABBR=DE82019300
   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.

9. 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." *National Academies Press*. Retrieved from http://www.nap.edu/catalog.php?record_id=13282
   Communicating Science and Engineering Data in the Information Age 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.

10. Pryor, G. ed. (2012). "Managing Research Data." *Facet Publishing*. http://www.facetpublishing.co.uk/title.php?id=7562
    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.

Citation Guides

Library Resource Guides on Data Citation

1. Cambridge. http://www.lib.cam.ac.uk/dataman/pages/citations.html
2. CDL. http://dcxl.cdlib.org/?p=233 and http://www.cdlib.org/services/uc3/dmp/citing.html
3. Minnesota. http://www.lib.umn.edu/datamanagement/cite
4. MIT. http://libraries.mit.edu/guides/subjects/data/access/citing.html
5. MSU. http://libguides.lib.msu.edu/citedata
6. Oregon. http://libweb.uoregon.edu/datamanagement/citingdata.html
7. Purdue. http://guides.lib.purdue.edu/datacitation
8. Toronto. http://datalib.chass.utoronto.ca/caq/citation.doc
9. UCambridge. http://www.lib.cam.ac.uk/dataman/pages/citations.html
10. UMinn. http://www.lib.umn.edu/datamanagement/cite
11. UVirginia. http://www2.lib.virginia.edu/brown/data/citing.html
12. UWM. http://www4.uwm.edu/libraries/AGSL/agsgis/find.cfm

Non-Library Guides to Data Citation

13. ANDS. “Data citation.” http://ands.org.au/ and http://www.ands.org.au/guides/data-citation-awareness.pdf
14. Argonne National Laboratory. “Argonne Premium Coal Samples Citation Form.” http://web.anl.gov/PCS/citation.html
15. DataCite. http://schema.datacite.org/meta/kernel-2.2/example/datacite-metadata-sample-v2.2.xml
16. DCC. http://www.dcc.ac.uk/resources/how-guides/cite-datasets. Overview here: http://www.dcc.ac.uk/webfm_send/295
17. Dryad. http://www.datadryad.org/using
18. EOL. http://eol.org/info/citing
19. GESIS Data Archive. http://www.gesis.org/en/services/data-analysis/data-archive-service/citation-of-research-data/
20. ICPSR. http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp
21. International Polar Ice Year. http://ipydis.org/data/citations.html.
22. NASA. http://history.nasa.gov/citeguide.html
23. NASA PDS. http://ppi.pds.nasa.gov/citations_policy.jsp
24. NOAA. http://www.ncdc.noaa.gov/paleo/citation.html
25. Pensoft. http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf
26. Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.org/citations
27. Statistic Canada. http://www.statcan.gc.ca/pub/12-591-x/2009001/steps-etapes-eng.htm
28. STD-DOI (German Science Foundation). http://dc110dmz.gfz-potsdam.de/contenido/std-doi/front_content.php?client=8&lang=7&idcat=1085&idart=182&m=&s=
29. UK Data Archive
    http://www.esds.ac.uk/doc/6654%5Cmrdoc%5CUKDA%5CUKDA_Study_6654_Information.htm
30. United States Department of Agriculture (2012). “Soil Data Access- Citation.” http://sdmdataaccess.nrcs.usda.gov/Citation.htm
31. USGS LP DAAC. https://lpdaac.usgs.gov/about/citing_lp_daac_and_data
32. Ball, A., Duke, M. (2011). “How to cite datasets and link to publications.” *DCC How-to Guides*. Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides/cite-datasets
    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.
33. Page, M., (1995). “A Brief Citation Guide for Internet Sources.” Retrieved from http://history.nasa.gov/citeguide.html

Citation Software and Repositories

1. ArXiv (Cornell). http://arxiv.org/
2. Australian National Data Service. http://www.ands.org.au/
3. Australian Research Collaborative Services. http://www.arcs.org.au/index.php/services/data-services
4. BGI (Beijing Genomics Institute) Cloud Computing. https://cloud.genomics.cn/
5. BMC_BL_Data_repositories (list). https://docs.google.com/spreadsheet/ccc?authkey=COmDvOUB&key=0Aok0Od_Hhd1XdEdiRXVCbDlFWk8wNW5FYlBBTndyaVE&hl=en_US&authkey=COmDvOUB#gid=0
6. Cambridge Crystallographic Data Centre. http://www.ccdc.cam.ac.uk/
7. Data.gov. www.data.gov
8. DATAPASS. http://www.icpsr.umich.edu/icpsrweb/content/DATAPASS/citations.html
9. DanBIF. http://www.danbif.dk/
10. Dataverse. http://thedata.org/
11. DTOL. https://sites.google.com/site/datatolproject/schema
12. Dryad. http://datadryad.org/
    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. http://www.dspace.com/en/inc/home.cfm
14. EBI. http://www.ebi.ac.uk/
15. ESDS. http://www.esds.ac.uk/international/. Video here: http://www.youtube.com/watch?v=NDrNHRjtd4g
16. EndNote. http://www.endnote.com
17. Figshare. http://figshare.com/
18. FISH.Link. http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/data-citation-and-linking
    Produced tools for converting and mapping freshwater biology data to linked data, while supporting semantic markup, attribution and provenance
19. Galaxy. http://galaxy.psu.edu/
20. GBIF. http://www.gbif.org/
21. GenBank. http://www.ncbi.nlm.nih.gov/genbank/
22. Giga Science (& British Library). http://www.gigasciencejournal.com/
23. ICPSR. http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp
24. INSPIRE SDI. http://www.intergraph.com/global/uk/government/INSPIRE.aspx. Long term preservation of here (PPT): http://inspire.jrc.ec.europa.eu/events/conferences/inspire_2010/presentations/55_pdf_presentation.pdf
25. International Virtual Observatory Alliance. http://www.ivoa.net/
26. LOCKSS. http://www.lockss.org/
27. Mendeley. http://www.mendeley.com
28. Mint (Molecular INTeraction Database). http://160.80.34.4/mint/Welcome.do
29. National Snow and Ice Data Center (NSIDC). http://nsidc.org/
30. NERC. http://ndg.badc.rl.ac.uk/
31. NGDA. http://www.ngda.org/
32. ORCID. http://about.orcid.org/
33. ORNL DAAC. http://daac.ornl.gov/
34. PANGEA. http://www.pangaea.de/
35. Polar Information Commons. http://www.polarcommons.org/ethics-and-norms-of-data-sharing.php
36. PDB (Protein Data Bank). http://www.rcsb.org/pdb/home/home.do
37. Publishing Network for Geoscientific & Environmental Data. http://www.pangaea.de/
38. RefWorks. http://www.refworks.com
39. SAEON. (policy) http://saeon.qsens.net/documentation/it-governance/policies-and-guidelines/data-policy-_stand-alone_.pdf/view
40. SAGECite. http://www.sagebase.org/.
    Produced a demonstrator citation service for network models, workflows and associated data in the Sage Commons, using a linked data approach.

41. SEAD (Sustainable Environment Actionable Data). http://sead-data.net/
42. SND. http://snd.gu.se/en
43. UnitProt (Universal Protein Resource Knowledgebase). http://www.uniprot.org/
44. Zotero. http://www.zotero.org

Conferences, Workshops, Symposia, Meetings

1. "Beyond the PDF." (2011) Retrieved from https://sites.google.com/site/beyondthepdf/

2. BRDI (2011). "Developing Data Attribution and Citation Practices and Standards." *An International Symposium and Workshop August 22-23, 2011, Berkeley, Ca*. Retrieved from http://sites.nationalacademies.org/PGA/brdi/PGA_064019

3. CLADDIER (2007). "Linking data and publications in the environmental sciences: CLADDIER project workshop." Retrieved from http://www.mendeley.com/research/linking-data-and-publications-in-the-environmental-sciences-claddier-project-workshop-chilworth-southampton-uk-15th-may-2007/

4. ANDS (Australian National Data service). "Data Citation Awareness." Retrieved from http://ands.org.au/guides/data-citation-awareness.html

5. Donnelly, M., Jones, S. (2011). "More with less: Collaborative trends in research data management." *PPT. Data management planning workshop, IDCC Conference, Bristol, England, December 5, 2011*. Retrieved from http://www.dcc.ac.uk/events/idcc11/workshops

6. European Science Foundation (2007). "Shared responsibilities in sharing research data: Policies and partnerships." Retrieved from http://www.knowledge-exchange.info/Default.aspx?ID=66&M=News&PID=177&NewsID=24

7. "IASSIST 2011-Data Science Professionals: A Global Community of Sharing." Retrieved from http://www.iassistdata.org/conferences/archive/2011

8. Kelly, M.C. (2008). "NISO thought leader meeting on research data." Memorandum. Retrieved from http://www.niso.org/topics/tl/NISOTLDataReportDraft.pdf

9. "Metadata for managing scientific research data." Webinar. August 22, 2012. Retrieved from http://www.niso.org/news/events/2012/dcmi/scientific_data/

10. Meeting with Ocean Science Journal Editors (2008). Retrieved from http://www.scor-int.org/Project_Summit_3/Data_Publication.pdf

11. National Center for Atmospheric Research (2012). "Bridging Data Lifecycles: Tracking Data Use via Data Citations Data Workshop." Retrieved from http://library.ucar.edu/data_workshop/

12. Harvard University (2011). "Principles of Data Citation, sponsored by Quantitative Social Science." Retrieved from http://projects.iq.harvard.edu/datacitation_workshop/

13. Workshop on Persistent Identifiers for the Social Sciences, sponsored by the IDSC of IZA/Gesis/RatSWD http://www.iza.org/conference_files/PeIdSS2011/viewProgram?conf_id=2013

Journals issues devoted to data

1. The Economist, "Data, data everywhere." February 27, 2010. Retrieved from http://www.economist.com/node/15557443

2. Nature, volume 455 (2008). “Special Issue: Big Data.” Retrieved from http://www.nature.com/nature/journal/v455/n7209/

3. Science, volume 331, 11 February 2011. Retrieved from http://www.sciencemag.org/content/331/6018.toc

Op-eds, Newsletters, Press Releases, Memorandums

1. Berman, F. (2010). “We Need a Research Data Census.” *Communications of the ACM Vol. 53 No. 12, Pages 39-41*. Retrieved from http://cacm.acm.org/magazines/2010/12/102121-we-need-a-research-data-census/fulltext
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.” *Published June 6, 2011 The Rockefeller University Press*. Retrieved from http://jem.rupress.org/content/early/2011/06/01/jem.20111061.full.pdf
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.” *FLUXNET Newsletter*. Retrieved from http://daac.ornl.gov/ornl_daac_citations_200812.pdf

4. “Datacite - Memorandum of Understanding” (2010). Retrieved from http://datacite.org/datacite_memo_understanding

5. Geoscience Information Society (2004). “Geoscience Information Society Newsletter.” Number 210, October 2004. Retrieved from http://www.geoinfo.org/GSIS_Newsletter/200410.pdf

6. Helly, J. New concepts of publication. Nature, 393, 1998. http://www.nature.com/nature/journal/v393/n6681/full/393107c0.html

7. Knecht L, Auld VA, McGhee M. (2001). “Changes in the Treatment of Chemical Data in MEDLINE® Citations.” *NLM Tech Bulletin*. Retrieved from http://www.nlm.nih.gov/pubs/techbull/nd01/nd01_mesh_chemical.html

8. Kolowich, S., (2011). “Killing Peer Review.” *Inside Higher Ed*. Retrieved from http://www.insidehighered.com/news/2011/07/19/debate_over_whether_social_web_sites_can_replace_peer_review

9. Lowry, R., Urban, E., & Pissierssens, P. (2009). “A New Approach to Data Publication in ocean sciences.” *Eos, Vol. 90, No.* 50. Retrieved from http://www.agu.org/pubs/crossref/2009/2009EO500004.shtml

10. Maunsell, J. (2010). “Announcement regarding supplemental material.” *The Journal of Neuroscience 11 August 2010, 30(32): 10599-10600.* Retrieved from http://www.jneurosci.org/content/30/32/10599.full
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.

11. NISO (2005). “NISO-Sponsored INFO URI Scheme Gets Thumbs Up from IETF Group.” Retrieved from http://www.niso.org/news/pr/view?item_key=4b8a9e2d84fe28e5559d725eb6acd6fd9b1eb53d

12. PR Newswire (2010). "Elsevier and PANGAEA Take Next Step in Connecting Research Articles to Data." *United Business Media*. Retrieved from http://www.prnewswire.com/news-releases/elsevier-and-pangaea-take-next-step-in-connecting-research-articles-to-data-99533624.html

13. Priem, J., Taraborelli, D., Groth, P., Neylon, C. (2010). "altmetrics: a manifesto." *Altmetrics*. Retrieved from http://altmetrics.org/manifesto/

Organizations, Committees

1. ANDS. http://www.ands.org.au/
2. ASIS&T SIG/MET. http://www.asis.org/SIG/met.html
3. CODATA, Task Force on Data Citation. http://www.codata.org/taskgroups/TGdatacitation/
4. Corporation for National Research Initiatives, http://www.cnri.reston.va.us/about_cnri.html
5. CSIR. http://www.csir.co.za/nre/ecosystems/Geoportal.html
6. DataCite. http://datacite.org
7. Data.gov. http://www.data.gov/
8. Dataverse. http://thedata.org/
9. Digital Curation Centre (DCC). http://www.dcc.ac.uk/
10. ESIP. http://www.esipfed.org/
11. FGDC. http://www.fgdc.gov/
12. Geoscience Information Society Task Force on citing geoscience data. http://www.geoinfo.org/TFGeosciData.htm
13. International Polar Year. http://classic.ipy.org/international/joint-committee/data-management.htm
14. JISC. http://www.jisc.ac.uk/aboutus.aspx
15. NDIPP. http://www.digitalpreservation.gov/
16. OAI. http://www.openarchives.org/
17. OSTI. http://www.osti.gov/
18. PARSE Insight. http://www.parse-insight.eu/project.php. Latest report here: http://www.parse-insight.eu/downloads/PARSE-Insight_D2-2_Roadmap.pdf
19. Research Data Canada. National Consultation on Access to Scientific Research Data (NCASRD), Canada. http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/ncasrd/
20. Southampton Data Management. http://www.southamptondata.org/
21. SAEON. http://www.saeon.ac.za/ Data policy here: http://data.saeon.ac.za/documentation/it-governance/policies-and-guidelines/data-policy-_stand-alone_.pdf/view
22. SURF (SURFShare). http://www.surffoundation.nl/en/themas/openonderzoek/Pages/Default.aspx
23. TIB. http://www.tib-hannover.de/en/
24. UKOLN. http://www.ukoln.ac.uk/

Papers

1. Aalbersberg, I. and Kahler, O. (2011). "Supporting Science through the Interoperability of Data and Articles." *D-Lib Magazine January/February 2011 Volume 17, Number ½.* Retrieved from http://www.dlib.org/dlib/january11/aalbersberg/01aalbersberg.html#3
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 from http://www.ijdc.net/index.php/ijdc/article/view/98/73
Early snapshot of CDL.

3. Acord, S., Harley, D. (in press). “Credit, time, and personality.” *New Media and Society*. Retrieved from http://nms-theme.ehumanities.nl/manuscript/credit-time-and-personality-acord-and-harley
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. Altman, M., Klass, G. M. (2005). “Current research in voting, elections, and technology.” *Social Science Computer Review Fall 2005 vol. 23 no. 3 269-273*. Retrieved from http://ssc.sagepub.com/content/23/3/269.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.

5. 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 from http://www.box.net/shared/d3cf8u0gtyml2nqq3u2f
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 (Altman et al., 2001). 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.

6. 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 from http://archivists.metapress.com/content/EU7252LHNRP7H188
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.

7. Altman, M. (2008). “A Fingerprint Method for Verification of Scientific Data.” *A Fingerprint Method for Verification of Scientific Data. : Springer-Verlag*. Retrieved from http://thedata.org/publications/fingerprint-method-verification-scientific-data
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.

8. Altman, M., Rogerson, K. (2008). “Open Research Questions on Information and Technology in Global and Domestic Politics – Beyond “E-.” *PS: Political Science & Politics, 41: pp835-837*. Retrieved from http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=2315604
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.

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

10. Altman, M., & Crabtree, J. (2011). "Using the SafeArchive System : TRAC-Based Auditing of LOCKSS." *Archiving 2011 (pp. 165-170*). Society for Imaging Science and Technology. Retrieved from http://www.imaging.org/IST/store/epub.cfm?abstrid=44591
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.

11. 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 from http://docs.lib.purdue.edu/iatul2010/conf/day1/1/
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.

12. Anderegg, W., Prall, J., Harold, J., Schneider, S. (2010). "Expert credibility in climate change." *Proceedings of the National Academy of Science.* Retrieved from http://www.pnas.org/content/early/2010/06/22/1003187107.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.

13. 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 Volume 5 Issue.* Retrieved from http://dl.acm.org/citation.cfm?id=1265746
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.

14. Autodesk Geospatial (2007). “Best Practice for Managing Geospatial Data.” Retrieved from http://www.gisperfect.com/res/AutocadMAP/best_practices.pdf
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

15. Ball, A., Duke, M. (2011). “Data Citation and Linking.” *Digital Curation Centre*. Retrieved from http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/data-citation-and-linking

16. Bernstein, H. J., Folk, M. J., Benger, W., Dougherty, M. T., Eliceiri, K. W. and Schnetter, E. (2011). “Communicating Scientific Data from the Present to the Future. Dowling College position paper.” Temporary URL: http://www.columbia.edu/~rb2568/rdlm/Bernstein_Dowling_RDLM2011.pdf

17. Bollen, J., Sompel, H. (2006). “An Architecture for the Aggregation and Analysis of Scholarly Usage Data.” *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries.* Retrieved from http://arxiv.org/abs/cs/0605113
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.

18. 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*. Retrieved from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0004803
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.

19. Bollen, J., Sompel, H. (2006). “Mapping the structure of science through usage.” *Scientometrics, Volume 69, Number 2, November 2006 , pp. 227-258(32)*. Retrieved from http://public.lanl.gov/herbertv/papers/Papers/2006/SCIENTObollen_map.pdf
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.

20. Bollen, J., Sompel, H., HagBerg, A., Chute, R. (2009). “A principal component analysis of 39 scientific impact measures.” *Cornell University Library*. Retrieved from http://arxiv.org/abs/0902.2183
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.

21. Bollen, J., Rodriguez, M., Sompel, H. (2006). “Journal Status.” *Scientometrics, volume 69, number 3, pp. 669-687, 2006.* Retrieved from http://arxiv.org/abs/cs.DL/0601030
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.

22. Bollen, J., Sompel, H., Smith, J., Luce, R. (2005). “Toward alternative metrics of journal impact: a comparison of download and citation data.*” Information Processing & Management Volume 41 Issue 6 Pagination 1419-1440.* Retrieved from http://arxiv.org/abs/cs.DL/0503007
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.

23. Bollen, J., Sompel, H., Rodriguez, M. (2008). “Towards Usage-based Impact Metrics: - First Results from the MESUR Project.” *Proceedings of the Joint Conference on Digital Libraries 2008*. Retrieved from http://arxiv.org/abs/0804.3791
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.

24. Bollen, J., 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 Volume 59 Issue 1, January 2008*. Retrieved from http://arxiv.org/pdf/cs.DL/0610154.pdf
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.

25. Borgman, C. (2011). "The conundrum of sharing research data." *Journal of the American Society for Information Science and Technology, pp. 1-40, 2011*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155
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.

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

27. Bose, R., Frew, J. (2005). "Lineage Retrieval for Scientific Data Processing." *ACM Computing Surveys, Volume 37, Issue 1, 2005)*. Retrieved from http://dl.acm.org/citation.cfm?id=1057978
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.

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

29. Brase, J., Farquhar, A., Gastl, A., Gruttemeier, H., Heijne, M., Heller, A., Hitson, B., Johnson, L., McMahon, B., Piguet, A., Rombouts, J., Sandfaer, M., & Sens, I. (2009). "Numeric Data: Citation Techniques and Integration with Text." Retrieved from http://www.icsti.org/IMG/pdf/Numeric_Data_FINAL_report.pdf
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.

30. Brase, J. (2004). "Using Digital Library Techniques- Registration of Scientific Primary Data." *Research and Advanced Technology for Digital Libraries 8th European Conference, ECDL 2004, Bath, UK, September 12-17, 2004. Proceedings.* Retrieved from http://www.springerlink.com/content/1pglbmjv95tqby9e/
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.

31. Brown, D., Welch, G., Cullingworth, C. (2005). "Archiving, management and preservation of Geospatial data." Retrieved from http://www.geoconnections.org/publications/policyDocs/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf
GeoConnections Policy Node created a working group that is used to identify issues and solutions related to long-term archiving and preservation of geospatial data. Geospatial data is produced by the government and private sector at an unprecedented rate. Au Yuen (2004) "…the real solution for digital preservation may lie less in technology and more in policy"

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

33. Buneman, P. Silvello, G (2010). "A Rule-Based Citation System for Structured and Evolving Datasets." *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*. Retrieved from http://sites.computer.org/debull/A10sept/buneman.pdf
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.

34. Callaghan, C., Donegan, S, Pepler, S. Thorley, M., Cunningham, N., Kirsch, P. 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). Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/208
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.

35. 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 (2002) Volume: 6, Issue: 2, Pages: 241.* Retrieved from http://www.mendeley.com/research/datasharing-datawithholding-genetics-life-sciences-results-national-survey-technology-transfer-officers-1/
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.

36. 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.* Retrieved from http://www.biomedcentral.com/1471-2105/10/S14/S2
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.

37. Cheney, J., Chiticariu, L., Tan,W.-T. (2009). "Provenance in databases: Why, where and how." *Foundations and Trends® in Databases: Vol. 1: No 4, pp 379-474.* Retrieved from http://www.nowpublishers.com/product.aspx?product=DBS&doi=1900000006
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.

38. CIESIN Columbia University (2005). "Data model for Manafing and preserving Geospatial Electronic Records." Retrieved from http://www.ciesin.columbia.edu/ger/DataModelV1_20050620.pdf

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.

39. Cole, F. (2008). "Taking "Data" (as a Topic): The Working Policies of Indifference, Purification and Differentiation." *Association for Information Systems Electronic Library*. Retrieved from http://aisel.aisnet.org/acis2008/79/
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.

40. 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 from www.daac.ornl.gov/DAAC/PI/bestprac.html#prac2
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

41. Costello, M. J. 2009. Motivating online publication of data. *Bioscience 59 (5): 418-427*. Retrieved from http://www.jstor.org/discover/10.1525/bio.2009.59.5.9?uid=3739912&uid=2&uid=4&uid=3739256&sid=55925848753
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.

42. Cragin, M. H., Palmer, C. L., Carlson, J.R., and Witt, M. (2010). "Data sharing, small science and institutional repositories." *Phil. Trans. R. Soc. A 13 September 2010 vol. 368 no. 1926 4023-4038.* Retrieved from http://rsta.royalsocietypublishing.org/content/368/1926/4023
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.

43. Crosas, M. (2011). “The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data.” *D-Lib Magazine January/February 2011 Volume 17, Number ½.* Retrieved from http://www.dlib.org/dlib/january11/crosas/01crosas.html
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.

44. DataCite (2011). “DataCite Metadata Scheme for the Publication and Citation of Research Data, Version 2.2, July 2011.” Retrieved from http://schema.datacite.org/meta/kernel-2.2/doc/DataCite-MetadataKernel_v2.2.pdf

45. 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.* Retrieved from http://www.ddialliance.org/node/100
Describes the best practices for metadata producers to provide end users with the resources for data discovery and dissemination. Citation not addressed

46. Downs, R. R., Chen, R.S. (2005). “Organizational needs for managing and preserving geospatial data and related electronic records.” *Data Science Journal Volume 4.* Retrieved from http://www.jstage.jst.go.jp/article/dsj/4/0/4_255/_article
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.

47. 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.* :1-22. Retrieved from http://dx.doi.org/10.1007/s12145-011-0083-6
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.

48. Fitzgerald, A. Pappalardo, K. (2007). “Building the infrastructure for data access and reuse in collaborative research.” Retrieved from http://eprints.qut.edu.au/8865/1/8865.pdf

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.

49. Freire, J., Koop, D., Santos, E., Silva, C. (2008). “Provenance for Computational Tasks: A Survey.” *Computing Science and Engineering, Vol 10, No 3, pp 11-21, 2008.* Retrieved from http://www.computer.org/portal/web/csdl/doi/10.1109/MCSE.2008.79
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.

50. Friends of the Chair Group on Integrated Economic Statistics (2007). “Session 3(c) – Dissemination standards (data and metadata), data exchange and revision policy.” http://www.bfs.admin.ch/bfs/portal/en/index/institutionen/statistikaemter_in/03/02.parsys.0021.downloadList.00211.DownloadFile.tmp/disseminationstandardsdataandmetadatadataexchangeandrevisionpolicyoecd3c.pdf

51. Fry, J., Houghton, J., Lockyer, S., Oppenheim, C., and Rasmussen, B., (2008). “Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes.” Retrieved from http://ie-repository.jisc.ac.uk/279/
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.

52. Gants, J. , Reinsel. D. (2010). “The digital universe decade – Are you ready?” Retrieved from 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

53. 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 from http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf
Mainly focusing on data growth.

54. Gibbs, H. (2007). “DISC-UK DataShare: State-of-the-art review.” Data Share project. Retrieved from http://www.disc-uk.org/docs/state-of-the-art-review.pdf

55. Gibbs, H. (2009). “Southampton data survey: Our experience and lessons learned.” University of Southampton. Retrieved from http://ie-repository.jisc.ac.uk/304/

56. Green, A., Macdonald, S., Rice, R. (2009). “Policy-making for Research Data in Repositories: A Guide.” *JISC funded DISC-UK Share Project*. Retrieved from http://www.disc-uk.org/docs/guide.pdf
Discusses citation briefly in the context of access and reuse of data. No survey done.

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

58. Greenberg, J. (2009). “Metadata Research Supporting the Dryad Data Repository.” *Cornell Univesity Library, eCommons@Cornell*. Retrieved from http://ecommons.library.cornell.edu/handle/1813/12247

Conference presentation. Citation not addressed

59. Hakala, J. (2010). “Persistent identifiers – an overview.” *The KIM Technology Watch Report*http://metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/
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.

60. Hamilton, E. (2007). “The impact of survey data: Measuring success.” *Journal of the American Society for Information Science and Technology Volume 58, Issue 2, pages 190–199, 15 January 2007.* Retrieved from http://onlinelibrary.wiley.com/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.

61. 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 from http://escholarship.org/uc/item/1xv148c8
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.

62. Heery, R. (2009). “Digital Repositories Roadmap Review: towards a vision for research and learning in 2013.” Retrieved from http://www.jisc.ac.uk/media/documents/themes/infoenvironment/reproadmapreviewfinal.doc
Addresses citation metrics

63. Helliwell, J. R. and McMahon, B. (2010). “The record of experimental science: Archiving data with literature.” Retrieved from http://iospress.metapress.com/content/f0765625774j4051/fulltext.pdf
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.

64. J. Helly, T. T. Elvins, D. Sutton, D. Martinez, S. Miller, S. Pickett, and A. M. Ellison (2002). “Controlled publication of digital scientific data.” *CACM 45(5).* Retrieved from http://dl.acm.org/citation.cfm?id=506222
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.

65. Hey, T., Trefethen, A. (2003). “The data deluge: An e-science perspective.” *From “Grid Computing – making the global infrastructure a reality”, Wiley*. Retrieved from http://eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf
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.

66. 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 from http://daac.ornl.gov/PI/BestPractices-2010.pdf
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.

67. 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.* Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2819144/
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.

68. Jones, S. (2012). “Developments in research funder data policy.” *International Journal of Digital Curation* 7(1). Retrieved from http://ijdc.net/index.php/ijdc/article/view/209/278
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 emphasises 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.

69. Kethers, S., Shen X., Treloar, A.E., Wilkinson, R. G.. (2010) “Discovering Australia’s Research Data.*” JCDL’10, June 21–25, 2010*. Retrieved from http://andrew.treloar.net/research/publications/jcdl2010/jcdl158-kethers.pdf
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.

70. King, G. (2007). “An Introduction to the Dataverse Network as an Infrastructure for Data Sharing.” *Sociological Methods & Research Volume 36 Number 2 November 2007 173-199*. Retrieved from http://gking.harvard.edu/gking/files/dvn.pdf

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.

71. 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 from http://escholarship.org/uc/item/9jw4964t#page-2
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.

72. Lane, M. (2008). "Data citation in the electronic environment." A white paper commissioned by GBIF. Retrieved from http://www.danbif.dk/Documents/gbif-documents/DataCitation-Lane2008.pdf

73. 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 Issue 2, Volume 6 | 2011.* Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/181/265
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.

74. Lyon, L. (2007). "Dealing with Data: Roles, Rights, Responsibilities and Relationships - Consultancy Report." Retrieved from http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/publications.html

75. Lyon, L., Rusbridge, C., Neilson, C. & Whyte, A. (2010). "Disciplinary approaches to sharing, curation, reuse and preservation." *DCC SCARP final report. JISC.* Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf

76. 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* Fall 2011 DOI:10.5062/F4CC0XMJ. Retrieved from http://www.istl.org/11-fall/article1.html
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.

77. Marcus, C. et. al. (2007). "Understanding research behaviors, information resources, and service needs of scientists and graduate students: A study by the university of Minnesota libraries." Retrieved from http://conservancy.umn.edu/bitstream/5546/1/Sciences_Assessment_Report_Final.pdf
Good for a general understanding of researcher behavior but little use for citation

78. Marical, L. Hemminger, B. (2010). “Scientific data repositories on the web: an intial survey.” *JASIST DOI: 10.1002/as*. Retrieved from http://www.ils.unc.edu/bmh/pubs/ScientificDataRepositories-JASIST-2010.pdf
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

79. 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 January/February 2011 Volume 17, Number ½*. Retrieved from http://www.dlib.org/dlib/january11/michener/01michener.html
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.

80. Moreau, L. (2010). “The Foundations for Provenance on the Web.” *Foundations and Trends® in Web Science: Vol. 2: No 2-3, pp 99-241*. Retrieved from http://eprints.soton.ac.uk/271691/1/survey.pdf
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.

81. Nelson, B. (2009). “Data sharing: Empty archives.*” Nature 461:160-163.*
Retrieved from http://www.nature.com/news/2009/090909/full/461160a.html

82. 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*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18445641
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.

83. Parsons, M.A., Duerr, R., Minster, J.-B. (2010). “Data citation and peer review.” *Eos, Transactions American Geophysical Union, 91(34), 297-298*. Retrieved from http://www.agu.org/pubs/crossref/2010/2010EO340001.shtml

84. Parsons, M., Bruin, T., Tomlinson, S., Campbell, H., Godoy, O., LeClert, J.,et al.(2009). “The State of Polar Data—the IPY Experience.” Retrieved from http://ipydis.org/documents/State_of_Polar_Data20100514_distribute.pdf

85. 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 Retrieved from http://www.agu.org/pubs/crossref/2010/2010EO340001.shtml
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.

86. Paton, N.W. (2008). “Managing and sharing experimental data: standards, tools and pitfalls.” *Biochemical Society Transactions 36 (1), 33-36*. Retrieved from http://www.mendeley.com/research/managing-and-sharing-experimental-data-standards-tools-and-pitfalls/
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.

87. Pepe, A., Mayernik, M., Borgman, C. (2010). “From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web.” *Published online XXX in Wiley InterScience DOI: 10.1002/asi.21263.* Retrieved from http://public.lanl.gov/herbertv/papers/Papers/2010/jasist_ore.pdf

88. Pepler, S. (2008). “Citation, location and deposition in discipline and institutional repositories.*” (CLADDIER): Final report. JISC*. Retrieved from http://ie-repository.jisc.ac.uk/223/1/final_report_v7-1.doc
Lots of theoretical work about citation, but no user surveys done.

89. Pepler, S. J., & O'Neil, K. (2008). “Preservation intent and collection identifiers: CLADDIER Project Report II.” Retrieved from http://epubs.cclrc.ac.uk/bitstream/2359/Report_II_PreservationIntentAndCompoundObjectIdentifiers-1.pdf
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

90. 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/
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.

91. Rodriguez, M., Bollen, J., Sompel, H. (2007). “A Practical Ontology for the Large-Scale Modeling of Scholarly Artifacts and their Usage.*” In Proceedings of the Joint Conference on Digital Libraries, Vancouver, June 2007.* Retrieved from http://public.lanl.gov/herbertv/papers/Papers/2007/JCDLrodriguez.pdf
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.

92. Rooyen, H. (2010). “Data on southern oceans now freely available.” Retrieved from http://ntww1.csir.co.za/plsql/ptl0002/PTL0002_PGE157_MEDIA_REL?MEDIA_RELEASE_NO=7523509
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.

93. 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, 2005, Volume 3652/2005, 128-138.* Retrieved from http://www.springerlink.com/content/2u3eng7kvt58t7v9/
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.

94. Schwartz, A., Pappas, C., Sandlow, J.(2010). “Data repositories for medical education research: issues and recommendations.” *Academic Medicine: May 2010 - Volume 85 - Issue 5 - pp 837-843*. Retrieved from http://journals.lww.com/academicmedicine/Abstract/2010/05000/Data_Repositories_for_Medical_Education_Research_.29.aspx
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.

95. 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, 0DOI: 10.1214/10-STS320. Retrieved from http://arxiv.org/pdf/1011.0810v1.pdf
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 issuesof 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.

96. Seeber, F. (2008). “Citations in supplementary information are invisible.*” Nature, 451(7181), 887.*
Retrieved from http://www.nature.com/nature/journal/v451/n7181/full/451887d.html

97. 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. Retrieved from
http://www.springerlink.com/index/10.1007/BF02628694
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.

98. Smit, E. (2011). "Avoiding a Digital Dark Age for data: why publishers should care about digital preservation." *Learned publishing* 24(1), 35-49. Retrieved from http://www.ingentaconnect.com/content/alpsp/lp/2011/00000024/00000001/art00007
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.

99. Smit, E. (2010). "Abelard and Héloise: Why Data and Publications Belong Together." *D-Lib Magazine January/February 2011 Volume 17, Number ½.* Retrieved from http://www.dlib.org/dlib/january11/smit/01smit.html
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.

100. Starr, J., &Gastl, A. (2011). "isCitedBy: A metadata scheme for DataCite." *D-Lib Magazine, 17(1/2). doi:10.1045/january2011-starr* Retrieved from http://www.dlib.org/dlib/january11/starr/01starr.html
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.

101. 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
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.

102. Takeda, K. (2010). "Institutional Data Management Blueprint Project." Retrieved from http://www.southamptondata.org/uploads/7/3/0/0/730051/idmbinitialfindingsreportv4.pdf. Talk here: http://www.eduserv.org.uk/newsandevents/events/eduserv-symposium-2011/the-institutional-data-management-blueprint-project Slides here: http://www.slideshare.net/eduserv/institutional-data-management-blueprint-7979017

103. Thessen, A., Patterson, D. (2011). "Data issues in the life sciences." *White paper*. Retrieved from http://dataconservancy.org/sites/default/files/Data%20Issues%20in%20the%20Life%20Sciences%20White%20Paper.pdf

104. Tilmes, C., Yesha, Y., Halem, M. (2011). "Distinguishing Provenance Equivalence of Earth Science Data." *Procedia Computer Science Volume 4, 2011, Pages 548–557*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1877050911001153
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

105. "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 from http://www.geoinfogeochem.org/sites/geoinfogeochem.org/files/Policy_GeochemDataPubl_v1.1_0.pdf

106. Vision, T.J. (2010). "Open data and the social contract of scientific publishing." *American Institute of Biological Sciences, 60(5), 330-331.* Retrieved from http://caliber.ucpress.net/doi/abs/10.1525/bio.2010.60.5.2

107. W3C. Incubator report (2010). Retrieved from http://www.w3.org/2005/Incubator/prov/XGR-prov-20101214/.
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

108. Walker, D., (2010). "The physics of complex systems in information and biology." *State University Of New York At Stony Brook, 2008, 123 pages.* Retrieved from http://gradworks.umi.com/33/86/3386263.html
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.

109. 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 Issue 1, Volume 3* . Retrieved from http://www.ijdc.net/index.php/ijdc/article/viewFile/67/46
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.

110. Waters, D. (2004). "Building on success, forging new ground: The question of sustainability." *First Monday, volume 9, number 5 (May 2004).* Retrieved from http://outreach.lib.uic.edu/www/issues/issue9_5/waters/index.html
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.

111. Wynholds, L. (2011). "Linking to scientific data: Identity problems of unruly and poorly bounded digital objects." *International Journal of Digital Curatio*n 6(1).Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/174
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.

Posters, Charts

1. BMC BL Data repositories
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
https://docs.google.com/spreadsheet/ccc?authkey=COmDvOUB&key=0Aok0Od_Hhd1XdEdiRXVCbDlFWk8wNW5FYlBBTndyaVE&hl=en_US&authkey=COmDvOUB#gid=0

2. 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 from
http://dataonedatacitations.wordpress.com/2010/09/13/dcc-poster-submission-data-citation-in -the-wild/

3. Newton, M. Mooney, H, Witt, M. "A Description of Data Citation Instructions in Style Guides." Retrieved from http://docs.lib.purdue.edu/lib_research/121/

4. NIH data sharing regulations/policy/guidance/ chart for NIH awards. Retrieved from grants.nih.gov/grants/policy/data_sharing/data_sharing_chart.doc

5. Pepe, A., Borgman, C. (nd) "Integrating scientific data into scholarly value chains." Retrieved from http://research.microsoft.com/en-us/events/ersymposium2009/integrating_scien_data_scholarly_value_chains.pdf

6. Piwowar,H. Chapman, W. (2007) "Examining the uses of shared data," Poster.
Retrieved from 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."
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 from http://vimeo.com/34338054.

3. Chavan, V. (2010). "Data Citation Mechanism and Services for primary biodiversity data." *Global Biodiversity Information Facility*. Retrieved from
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 from http://info.nfais.org/info/ChenDownsNov10.pdf

5. Coggins, J. (2009). "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*. Retrieved from http://www.ukoln.ac.uk/events/ukrds-2009/programme/

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

7. ESDS International. "Citing Data." Retrieved from http://youtu.be/NDrNHRjtd4g

8. IASSIST. (2011). "Session B2: The IASSIST SIGDC Presents: Perspectives on Data Citation (Wed, 2011-06-01); Session C2: DataCite - Making Data Citable (Wed, 2011-06-01)." Retrieved from http://www.iassistdata.org/conferences/archive/2011

9. King, G. (2007). "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing." Retrieved from http://www.youtube.com/watch?v=fgn6dmfsZ_M

10. Primary Data at WDC Climate (WDCC *)." MPG eScience Seminar: Persistent Identifier Garching*. Retrieved from 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.iiaweb.com

12. McMahon, B. (2010). "Integrating Data with Publications: Greater Interactivity and Challenges for Long-Term Preservation of the Scientific Record." *Crystallography Journals Online.* Retrieved from 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 from http://libraries.mit.edu/guides/subjects/data-management/Managing%20Research%20Data%20101.pdf

14. Mooney, H. (2010). "Data Reference In Depth: Citation." *IASSIST 2010 Conference- June 4, 2010.* Retrieved from 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." 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 from http://river-valley.tv/conferences/ebooks-econtent-2010

Reports

1. Australian Government (2007). "Towards the Australian Data Commons." Proposal for the ANDS. Retrieved from http://www.pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf

2. 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 from http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

3. Bohn, R., Short, J. (2009). "How Much Information? 2009 Global Information Industry Center Report on American Consumers." Retrieved from http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf
Focusing on information consumption.

4. CCSDS (Consultative Committee for Space Data Systems) (2002). *"Reference Model for an Open Archival Information System." (OAIS) CCSDS 650.0-B-1 Issue 1. Washington, DC: CCSDS Secretariat*. Retrieved from http://public.ccsds.org/publications/archive/650x0b1.PDF

5. 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

6. CSIR (2011). "Geoportal offers larger spatial data seta at the click of a button." *Ecosystems. Earth observation.* Retrieved from http://www.csir.co.za/nre/ecosystems/Geoportal.html
It's 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 whatis explained.

7. EDUCAUSE (2006). "IT Engagement in Research: A Baseline Study." Retrieved from http://www.educause.edu/ECAR/ITEngagementinResearchABaselin/158595
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

8. High Level Expert Group on Scientific Data (2010). "Riding the Wave: How Europe Can Gain from the Rising Tide of Scientific Data. European Commission." Retrieved from http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf
Support for scientific e-infrastructure that supports seamless access, use, re-use, and trust of data.

9. International Association of Scientific, Technical , and Medical Publishers. (2007). "Brussels Declaration of 2007." http://www.stm-assoc.org/brussels-declaration/

10. 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 from http://www.arl.org/bm~doc/ARL_EScience_final.pdf.
Article relates to eScience and libraries. Citation was not mentioned.

11. National Aeronautics and Space Administration (2002). "Reference Model for an Open Archival Information System." *2002 Recommendation for Space Data System Standards, Consultative Committee for Space Data*

Systems Secretariat, Program Integration Division (Code M-3), National Aeronautics and Space Administration. Retrieved from http://public.ccsds.org/publications/archive/650x0b1.pdf on 4 October 2006.

12. 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 from: http://digitalpreservation.gov/documents/NDSA_ResponseToOSTP.pdf

13. National Information Standards Organization (NISO), National Federation of Advanced Information Services (NFAIS) (2010). “Roundtable on Best Practices for Supplemental Journal Article Materials.” Retrieved from http://iassist-sigdc.googlegroups.com/attach/7186703f23266e75/RP-15-201x+Suppl_BWG_draft_for_comments.pdf?view=1&part=2
An informal survey of large publishers' practices in accepting, processing, publishing, and preserving supplementary materials was conducted in October 2009 by Alexander (Sasha) Schwarzman, AGU Information Systems Analyst. The publishers were contacted via CrossRef TWG and eXtyles listservs. While all of the publishers surveyed were distributing these types of materials, there was little consistency in how they were handled. There was consensus in the view that all supplemental materials should be peer-reviewed, but not necessarily about the rigor of that review. The size and scope of the supporting materials was an issue, as well as if and where those materials reside online. Publishers generally responded that supplemental materials did not go through the same production processes, such as editing, layout, consistent markup, etc. While ensuring that the supporting data remained intact and unchanged, this lack of production management could lead to problems when a publisher wants to archive the information or migrate it to a future system.
Shorter version here:
http://www.niso.org/apps/group_public/download.php/3708/NFAIS_NISO_Supp_Materials_Meeting_Summary_Report_rev.pdf
Supplemental information regarding survey here:
http://www.agu.org/dtd/Presentations/sup-mat/10.3789_isqv22n3.2010.05.pdf
http://www.agu.org/dtd/Presentations/sup-mat/sup-mat.pdf

14. National Science and Technology Council, Interagency Working Group on Digital Data (2009). “Harnessing the Power of Digital Data for Science and Society.” Retrieved from http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

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

16. National Science Foundation (2007). “Cyberinfrastructure Vision for 21st Century Discovery.” Retrieved from http://www.nsf.gov/pubs/2007/nsf0728/.

17. National Science Foundation (2011). “Digital research data sharing and management.” Retrieved from http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf
Recommendations are organized under four areas: commitment to sharing; reproducibility; education, training, and workforce development; and longevity and sustainability.

18. National Science Foundation (2011). “Division of Ocean Sciences Sample and Data Policy.” Retrieved from http://www.nsf.gov/pubs/2011/nsf11060/nsf11060.pdf

19. National spatial data infrastructure (2003). “Managing Historical Geospatial Data Records.” Retrieved from http://www.fgdc.gov/library/factsheets/documents/histdata.pdf
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.

20. Organization for Co-operation and Development (2007). “OECD Principles and Guidelines for Access to Research Data from Public Funding.” Retrieved from http://www.oecd.org/dataoecd/9/61/38500813.pdf

21. Research Information Network (2008). “To share or not to share.” Retrieved from http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs

22. UNESCO (2008). “SCOR/IODE Workshop on Data Publishing.” *IOC Workshop Report No. 207.* Retrieved from http://www.scor-int.org/Publications/wr207.pdf

Standards

1. DublinCore. http://dublincore.org/

2. IEEE. http://standards.ieee.org/

3. International Organization for Standardization (2010). ***ISO 690:2010. Information and documentation: Guidelines for bibliographic references and citations to information resources***. Geneva, Switzerland: Author.

4. ISO Standards. Retrieved from http://www.iso.org/iso/catalogue_tc.htm

5. ISO. 2003. ISO Standard 14721:2003, Space Data and Information Transfer Systems—A

6. Reference Model for An Open Archival Information System (OAIS). International Organization for Standardization. Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683

7. ISO. 2005. ISO Standard19110:200, Geographical Information- Methodology for Feature Cataloguing http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39965

8. National Information Standards Organization, & American National Standards Institute. (2005). *Bibliographic references*. Bethesda, MD: NISO Press. http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key:ustring:iso-8859-1=87775a75d6ea19921a41d75b2fb012b0d6339b3a

9. NISO. http://www.niso.org/standards/

Surveys & Studies

1. Arovelius et al. (2006). “SLU Pilot study: Digital preservation of research materials.” Retrieved from http://dspace.slu.se:8080/dspace/index.jsp
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 discusses three databases that were tested in the study.

2. Banach, M., Li, Y. (2011). “Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries.” Retrieved from http://works.bepress.com/meghan_banach/5/
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.

3. Beagrie, N., Chruszcz, J. and Lavoie, B., (2008). “Keeping Research Data Safe: a cost model and guidance for UK Universities.” *Joint Information Systems Committee 2008.* Retrieved from http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx
Costs associated with managing data.

4. Beagrie, N., Lavoie, B., Woollard, M. (2010). “Keeping Research Data Safe 2.” Retrieved from http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf

5. Beagrie, N., Beagrie, R., Rowlands, I. (2009). “Research data preservation and access: the views of researchers.” *Ariadne* 60. Retrieved from http://www.ariadne.ac.uk/issue60/beagrie-et-al/
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.

6. 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 from http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP-Synthesis.pdf

7. 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.” *University of California, Berkeley: Center for Studies in Higher Education.* Retrieved from http://escholarship.org/uc/cshe_fsc
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.

8. Johnston, L. (2010). “User-needs assessment of the research cyberinfrastructure for the 21st century.*” Perdue University*. Retrieved from http://docs.lib.purdue.edu/iatul2010/conf/day1/5/
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

9. Lyman, Varian (2003). “How Much Information 2003.” Retrieved from http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf
Focusing on information production.

10. Maron, Nancy L. & Smith, Kirby (2008). “Current models of digital scholarly communication: Results of an investigation conducted by Ithaka for the Association of Research Libraries.” *Association of Research Libraries*. Retrieved from http://www.arl.org/bm~doc/current-models-report.pdf

11. 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 from http://staff.lib.msu.edu/mooneyh/myresearch/HMooney_Citingdatasources_preprint.pdf
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.

12. Pienta, A., Alter, G., Lyle, J. (2010). "The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data." Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/78307
    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.

13. Piwowar, H., Chapman, W. (2010). "Public sharing of research datasets: A pilot study of associations." ***148-156. In Journal of Informetrics 4 (2).*** Retrieved from http://www.sois.uwm.edu/MetricsPreCon/documentation/Piwowar_Chapman_Sharing.pdf
    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 highimpact 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.

14. Piwowar, H., Chapman, W. (2008). A review of the journal policies for sharing research data. In *ELPUB*. Retrieved from http://ocs.library.utoronto.ca//index.php/Elpub/2008/paper/view/684
    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. Pinowar, H. Day, R. Fridsma, D. (2007) "Sharing detailed research data is associated with increased citation rate." http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000308
    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.

16. Polydoratou, P. (2007). "Use of digital repositories by chemistry researchers: Results of a survey." *Program: Electronic library and information systems, 41, pp386–399*.Retrieved from http://www.emeraldinsight.com/journals.htm?articleid=1630903
    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, favourably 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 recognise 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.

17. 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 from http://hdl.handle.net/1802/6053
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.

18. Research Information Network (2011). “Data Centers: their use, value, and impact.” Retrieved from http://www.rin.ac.uk/our-work/data-management-and-curation/benefits-research-data-centres

19. *Research information network.*, (2011). “Information handling in collaborative research: an exploration of five case studies.” Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/collaborative-research-case-studies
The case studies focus on the behaviours 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 behaviours 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.

20. *Research information network.* (2011). “Physical Sciences Case studies: information use and discovery.” Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/physical-sciences-case-studies-use-and-discovery-
This project focused on the behaviours 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 behaviours 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.

21. Research information network. (2011). “Reinventing research? Information practices in the humanities.*”* Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities
This project focuses on the behaviours and needs of researchers working in a number of subject or disciplinary areas in the humanities. They follow the first round of case studies in the life sciences.

22. Research information network (2011). “The value of libraries for research and researchers.” Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/value-libraries-research-and-researchers
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.

23. Shaon, A., Woolf, A. (2010). “Long-term preservation for INSPIRE: a metadata framework and geo-portal implementation.” Retrieved from inspire.jrc.ec.europa.eu/events/conferences/inspire_2010/abstracts/55.doc
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.

24. Simmhan, Y., Plale, B., Gannon, D. (2005). “A survey of data provenance in e-science.” *ACM SIGMOD Vol 34, No 3, 2005*. Retrieved from http://pti.iu.edu/sites/default/files/simmhanSIGMODrecord05.pdf
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.

25. Soehner, C., Steeves, C., Ward, J. (2010). “e-Science and data support services: a survey of ARL members.” Retrieved from http://www.arl.org/bm~doc/escience_report2010.pdf
Surveyed users, but didn’t ask about citation.

26. Sukovic, S. (2009). “References to e-texts in academic publications.” *Journal of Documentation, Vol. 65 Iss: 6, pp.997 – 1015.* Retrieved from http://www.mendeley.com/research/references-to-etexts-in-academic-publications/
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.

27. 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*. Retrieved from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0021101
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.

28. 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 Med. 2010 Aug;12(8):486-95*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20535021
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.

29. Waaijers, L. and Van der Graaf, M. (2011). “Quality of Research Data, an Operational Approach.” *D-Lib Magazine January/February 2011 Volume 17, Number ½.* Retrieved from http://www.dlib.org/dlib/january11/waaijers/01waaijers.html
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.

30. Wolf, A., Simpson, M., Salo, D., Flee, D., Cheetham, J., Barton, B. (2009). “Summary Report of the Research Data Management Study Group.” Retrieved from http://minds.wisconsin.edu/handle/1793/34859
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). http://www.cru.uea.ac.uk/cru/projects/acrid/
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. http://www.alliancepermanentaccess.org/
3. CENS. http://research.cens.ucla.edu/
4. CIRES. http://cires.colorado.edu/
5. CMIP5. http://cmip-pcmdi.llnl.gov/cmip5/
6. CrossRef. http://www.crossref.org/
7. Datacurate. http://www.datacurate.com/
8. Data Interactive Publications. https://sites.google.com/site/datainteractivepublications/
9. Data Seal of Approval (Ball & Duke, UKLON). http://www.datasealofapproval.org/?q=node/66
10. Dataverse Network Citation Standard, http://thedata.org/citation
11. DigCurV (2010). “Digital Curator Vocational Education Europe.” http://www.digcur-education.org/
12. e-Bank. http://www.ukoln.ac.uk/projects/ebank-uk/data-citation/
13. Elsevier. www.elsevier.com

14. ESDS (Economic and Social Data Service), http://www.esds.ac.uk/international/news/news.asp#21sep11
15. EZID. http://n2t.net/ezid
16. Federal Register. https://www.federalregister.gov/articles/2011/11/04/2011-28621/request-for-information-public-access-to-digital-data-resulting-from-federally-funded-scientific
17. ICPSR (*Interuniversity Consortium for Political and Social Research*)- Data Citations http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.js
18. MMI. http://marinemetadata.org/
19. NERC. http://www.nerc.ac.uk/research/sites/data/
20. OakLawProject. http://www.oaklaw.qut.edu.au/
21. OECD (Organisation for Economic Co-operation and Development) (T. Green), http://www.oecd.org/home/0,3675,en_2649_201185_1_1_1_1,00.html
22. Openwetware. http://openwetware.org/wiki/Main_Page
23. Plos. http://www.plos.org/
24. SPQR. http://spqr.cerch.kcl.ac.uk/ (Supporting Productive Queries for Research) trialled 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.
25. STFC. http://www.stfc.ac.uk/e-Science/default.aspx
26. STM (International Association of Scientific, Technical & Medical Publishers), http://www.stm-assoc.org/about-the-association/
27. Thompson, http://researchanalytics.thomsonreuters.com/solutions/researcherid/
28. UC3DCXL. http://dcxl.cdlib.org/?page_id=11
29. Webtracks, http://www.stfc.ac.uk/e-Science/projects/medium-term/metadata/webtracks/22422.aspx 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.
30. XYZ Project, http://projectxyz.wordpress.com/
Developed tools and an exemplar workflow for co-ordinating the deposition of data in archive with the review and publication of an associated paper.

Appendix C: Additional Resources Not Yet Incorporated in Foregoing Analysis and Categorization

Title: Recommended practices for citation of data published through the GBIF network.

Author(s): Chavan, V.

Publisher(s): GBIF Secretariat, 2012

Abstract: 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.

Bibliographic citation: GBIF (2012). Recommended practices for citation of the data published through the GBIF Network. Version 1.0 (Authored by Vishwas Chavan), Copenhagen: Global Biodiversity Information Facility. Pp.12, ISBN: 87-92020-36-4. Accessible at http://links.gbif.org/gbif_best_practice_data_citation_en_v1

Rights: This document is licensed under a Creative Commons Attribution 3.0 Unported License

Rights Holder: GBIF Secretariat

Download: http://www.gbif.org/orc/?doc_id=4659&l=en

Title: The Anatomy of a Data Citation: Discovery, Reuse, and Credit

Authors: Hailey Mooney, Michigan State University; Mark P. Newton, Columbia University

Publisher: Pacific University Library

Abstract: INTRODUCTION Data citation should be a necessary corollary of data publication and reuse. Many researchers are reluctant to share their data, yet they are increasingly encouraged to do just that. Reward structures must be in place to encourage data publication, and citation is the appropriate tool for scholarly acknowledgment. Data citation also allows for the identification, retrieval, replication, and verification of data underlying published studies.
METHODS This study examines author behavior and sources of instruction in disciplinary and cultural norms for writing style and citation via a content analysis of journal articles, author instructions, style manuals, and data publishers. Instances of data citation are benchmarked against a Data Citation Adequacy Index.
RESULTS Roughly half of journals point toward a style manual that addresses data citation, but the majority

of journal articles failed to include an adequate citation to data used in secondary analysis studies. DISCUSSION Full citation of data is not currently a normative behavior in scholarly writing. Multiplicity of data types and lack of awareness regarding existing standards contribute to the problem. CONCLUSION 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.

Recommended Citation: Mooney, H, Newton, MP. (2012). The Anatomy of a Data Citation: Discovery, Reuse, and Credit. Journal of Librarianship and Scholarly Communication 1(1):eP1035.
Available at: http://jlsc-pub.org/jlsc/vol1/iss1/6
Mooney, H, Newton, MP. (2012). The Anatomy of a Data Citation: Discovery, Reuse, and Credit. Journal of Librarianship and Scholarly Communication 1(1):eP1035. Available at: http://jlsc-pub.org/jlsc/vol1/iss1/6