Canadian Association of Research Libraries

Association des bibliothèques de recherche du Canada

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

March 2010

By Kathleen Shearer and Diego Argáez

Canadian Association of Research Libraries (CARL)

On behalf of the CARL Data Management Sub-Committee:

Marnie Swanson (Chair, University of Victoria)

Pam Bjornson (NRC-CISTI)

Lynn Copeland (Simon Fraser University)

Michelle Edwards (University of Guelph)

Bernie Gloyn (Observer, Statistics Canada)

Margaret Haines (Carleton University)

Janine Schmidt (McGill University)

Kathleen Shearer (Research Associate, CARL)

Diego Argáez (Secretary, CARL)

Pavillon Morisset Hall,
65 rue University Street, Suite 239,
Ottawa Ontario K1N 9A5

613.562.5385 613.562.5195

carladm@uottawa.ca www.carl-abrc.ca

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

Table of Contents

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

Introduction

To paraphrase John Wilbanks, Vice President of Science Commons, 'our capacity to measure, store, analyze, and visualize data is becoming the new reality to which research will have to adapt'[1]. This is true not only for the sciences. Indeed, research data in all fields are increasingly being recognized as assets that have value beyond the original purpose for which they were created.

For the most part, research data are born digital, and stored and managed electronically, making them easy to share, replicate, and combine with other data. In 2006, a the results of a national consultation on research data in Canada envisioned "a new world of research and a whole new world."[2] The report goes on to argue that when databases are combined within and between disciplines and countries, fundamental leaps in knowledge can occur that transform our understanding of life, the world and the universe. In order to foster this 'new world', data must be created and maintained in a manner consistent with the goal of long-term preservation. This involves active management throughout the life cycle of the data, beginning at the time they are first envisioned. Adding to the complexity, there are wide variations in the use patterns, types, and formats of research data across disciplines.

In this burgeoning era of 'data intensive scholarship', roles and responsibilities have not yet been well established and there is an opportunity for libraries to develop unique and valuable services. Libraries, in many ways, are well positioned to support research data stewardship. They are already recognized on campus for preserving and providing access to other types of content; and they have strong links with the disciplinary communities. However, data also present special challenges. Research data are very diverse in both format and how they are used, and cannot be treated as a single document 'type'. In addition, 'data collections'[3] which have use beyond the original research project for which the data were created are community resources. This can make it difficult to demonstrate that library data services are directly benefiting the researchers at a given institution.

This document presents the results of a review of novel opportunities for libraries in the area of research data services. The activities were identified through a review of the literature and a scan of projects being undertaken at libraries and other institutions worldwide. For the purpose of this report, research data services have been organized into five distinct areas (although it should be noted that there are significant overlaps between them): awareness and advocacy; support and training; access and discovery; archiving and preservation; and virtual research environments. Each section contains a general description of the area accompanied by a number of examples. The examples are not meant to be comprehensive account of existing projects, but rather to highlight the range of possibilities available.

[1] Wilbanks, J. pg. 209
[2] Strong, et al. pg. 1
[3] The National Science Foundation characterizes reference data collections as "intended to serve large segments of the scientific and education community. Characteristic features of this category of digital collections are a broad scope and a diverse set of user communities including scientists, students, and educators from a wide variety of disciplinary, institutional, and geographical settings." See: National Science Foundation, Appendix D.

Awareness and Advocacy

Numerous consultations and studies conducted over the last decade have discussed the need for greater awareness of data stewardship with the various stakeholder communities. Libraries are well positioned to take on this role, especially in promoting awareness with researchers and university administrators.

A range of outreach activities can be undertaken by libraries to promote data stewardship. These include talking to stakeholders and researchers on campus; passing out literature; organizing conferences with data stewardship as a theme; and identifying academic staff to act as champions within departments. Because the issues and challenges of data stewardship can be very discipline specific, advocacy programs are most effective when they are tailored for a specific community. Messages should also be clear and consistent, and consideration should be given to the cultural issues across different disciplines that might act as barriers to widespread stewardship of research data.

Examples:

E-Science Talking Points for ARL Deans and Directors
www.arl.org/bm~doc/e-science-talking-points.pdf
This document was developed for ARL Deans and Directors to help them better understand e-science and cyberinfrastructure. The document defines key terms and describes some of the issues surrounding data. It also talks about relevant data policies (mainly in the US) and roles for libraries in evolving e-science environments.

CARL Data Management Awareness Toolkit
www.carl-abrc.ca/new/pdf/data_toolkit_communique_jan2010.pdf
The *Research Data: Unseen Opportunities* Awareness Toolkit was developed to enable library directors to raise awareness of the issues of data management with administrators and researchers on campus. Data are valuable assets that in some cases have an unlimited potential for reuse. The Awareness Toolkit underscores the need to ensure that research data are managed throughout the data lifecycle so that they are understandable and usable. *Research Data: Unseen Opportunities* provides readers with a general understanding of the current state of research data in Canada and internationally. It is organized into seven sections: The Big Picture, Major Benefits of Data Management, Current Context, Case Studies, Gaps in Data Stewardship in Canada, Data Management Policies in Canada, Responses to Faculty/Administrative Concerns, and What Can Be Done on Campus?

Data Audit at Oxford University
www.ict.ox.ac.uk/odit/projects/digitalrepository/
Data audits are another way of improving awareness of the issues with researchers. Auditors gather information about the data assets that currently exist: where they are located; how they have been managed; which need to be maintained in the long-term; and whether current practices put the data at risk. A Data Audit Framework was developed by the European Union and has been used by Oxford University Library (and others) to learn more about researchers' data practices and to capture their user requirements for services to support their data management.

Various Publications
There are a number of recently published articles that promote the idea of data stewardship and data sharing. These articles speak directly to the concerns of researchers and are valuable tools for educating faculty about the importance of data sharing and stewardship. A few of these publications are listed below:

1. Special section on Data Sharing. Nature, v. 461, no. 7261, 10 September 2009

2. *Data publication: towards a database of everything*. Vincent S Smith; BMC Research Notes v. 2, http://www.biomedcentral.com/1756-0500/2/113
3. *Data Sharing: The Next Generation*. Declan Butler; Nature v. 446, p. 10-11, March 1, 2007: www.nature.com/nature/journal/v446/n7131/full/446010b.html

Training and Support

In addition to advocacy and awareness activities, there is a real requirement for training and skills development in the research community. Researchers need advice concerning data management plans, technical standards, data cataloguing, metadata standards and processes, and preservation management. A Gap Analysis published by the Canadian Research Data Strategy Working Group in 2008, found that researchers rarely have the skills required to appropriately manage their data. The situation is similar in other countries. A recent UK survey, for example, found that "although there are disciplinary variances, data management competence of researchers is generally poor."[4]

Right now, there aren't a lot of opportunities in Canada for researchers to acquire data management training. The Research Data Strategy Working Group lead by CISTI (Canada Institute for Scientific and Technical Information) is in the process of developing a research data course, and there are some discipline specific opportunities for data training. Libraries, that already report they are being approached by researchers for advice and help, could fill in the gaps by providing access to training resources. Various strategies can be employed to improve researchers' skill levels, including the development of "how-to" guides and web-pages, training tutorials, and data management courses. It should be noted that because of the diverse nature of research data, higher level training resources should be developed with a specific disciplinary focus.

Examples:

MIT Libraries: Manage Your Data
http://libraries.mit.edu/guides/subjects/data-management/index.html
MIT has developed a very comprehensive website that provides guidance for researchers about managing their research data. The home page states, "Managing your data before you begin your research and throughout its life cycle is essential to ensure its current usability and long-run preservation and access." The site contains a "Data Management Checklist" and offers advice for developing data management plans. Other sections of the website include: Why Manage Your Data?, What is Data?, Evaluate Your Data Needs, Documentation and Metadata, File Formats, Organizing Your Files, Backups and Security, Sharing Your Data, Citing Data, Data Integration, Funding Requirements, Ethical and Legal Issues.

Australian National Data Service Guides
http://ands.org.au/guides/index.html
These guides provide information about some fundamental issues in data-intensive research and research data management. They are guidelines and checklists to inform and broaden the range of possibilities for researchers, data managers, and research organisations. The guides currently available through the ANDS site are: Copyright and data, Creative commons and data, Data citation, Data management planning, File formats, Identify My Data, Metadata, Persistent identifiers, Register My Data, Research Data Australia, Research data policy and the Australian Code for the Responsible Conduct of Research, Scholarly communications, and Storage.

[4] Lyon, L. pg.54

Research Data Management Forum: Digital Curation Centre
www.dcc.ac.uk/events/data-forum-2010-march/
The Research Data Management Forums were created in response to the challenges posed by the new and rapidly evolving digital environment. The underlying principle is to promote the exchange of experience and expertise in the management and curation of research data output as a means of improving its quality, reliability, processing, management and accessibility. To date, there have been 5 forums organized since 2007. Each forum selects a strategic issue to provide a framework for its activities whilst remaining anchored in the practical concerns of delivering data stewardship services.

IGDM02 - Data Management Course at the Australian National University
http://ilp.anu.edu.au/dm/
The course is based on the Data Management Manual developed by the Australian National University (ANU). It provides an introduction to Data Management. Participants learn what digital research data is and why it needs to be managed. The benefits of data management are discussed, as well as funding agency policies relating to data. Various DM concepts are discussed in detail, such as backups, version control, metadata, and archiving. During the course participants will also develop a draft Data Management Plan for their current research work. In addition, issues relating to data are explained and various methods for solving these problems with software tools and ANU services are described.

Managing and Sharing Data: A Best Practices Guide for Researchers.
www.data-archive.ac.uk/news/publications/managingsharing.pdf
Published by the UK Data Archive, the information guide is designed to help researchers and data managers across all research disciplines and research environments make sure that research data are of the highest quality and have the greatest potential for long-term re-use. From the forward, "Good data management is the foundation for good research. If data are properly organised and preserved, and their accuracy and integrity is controlled at all times, the result is high quality data, efficient research and the saving of time and resources. Researchers themselves benefit greatly from properly managing their research data. Data management should be planned from the start of research. If it becomes part of standard research practice, then it need not incur additional time or costs."

Discovery and Access

Although research data often have value beyond the purpose for which they were originally created, most data are still not shared or made publicly available. According to the 2008 Gap Analysis, "Much of the research data being produced today is hard to access by other Canadian research communities, and is often not ideally structured to be as useful or as open as possible, even within the discipline for which it is being constructed."[5]

Metadata are particularly important for ensuring that data can are understood by others and re-used. The National Science Foundation says, "To make data usable, it is necessary to preserve adequate documentation relating to the content, structure, context, and source (e.g., experimental parameters and environmental conditions) of the data collection – collectively called *metadata*. Ideally, the metadata are a record of everything that might be of interest to another researcher."[6]

[5] Research Data Strategy Working Group. pg. 13
[6] *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century.* pg. 20

Libraries could be instrumental in developing services aimed at improving access to data sets produced on campus- or elsewhere. These services might involve cataloguing and linking to diverse datasets held elsewhere. Or, more comprehensively, service could entail a range of data curation activities.

Examples:

Register My Data with the Australian National Data Service
http://www.ands.org.au/services/register-my-data.html
The ANDS Register My Data service allows researchers to register collections of research data. Descriptions of registered collections are published in a number of discovery services and e-research applications. Register My Data assists researchers and research organizations to publicize their research data collections, enabling verification of research findings and re-use of valuable research materials. ANDS aims to contribute to a global research data commons.

The ANDS Register My Data service is intended to be a comprehensive registry of research done in Australia. It registers the description of the data collection and endeavors to ensure that registered collections feature prominently in search engine results. In order to maintain the currency of the information, the Register My Data service supports a number of dynamic exchange and harvesting protocols to automate communication with managed data environments. Manual descriptions and registration are also supported.

Distributed Data Curation Center (D2C2) at Purdue University Libraries
http://d2c2.lib.purdue.edu/
The aim of the D2C2 is to address curation issues and work on problems related to unorganized, disparate, heterogeneous and distributed data, data workflow and environments. It works closely with the efforts of other agencies, centers, and groups which are doing related work so that practices and standards can be shared, reviewed and evaluated. The majority of projects involve working in partnership with domain scientists and information technologists to address the real world data needs of a research community. The scope of projects typically span across multiple areas including: data management, discovery and dissemination of data, data description and structure, social aspects, such as creating community based curation profiles and policies, and technology, such as designing data repository frameworks and creating software tools to add value and functionality to data sets.

ODESI at the Ontario Council of University Libraries
http://odesi.ca
<odesi> is a web-based tool for data exploration, extraction and analysis. <odesi> is a jointly funded project between the Ontario Council of University Libraries (OCUL) and OntarioBuys that provides university researchers with unprecedented access to a significant number of datasets. The project targets Statistics Canada datasets, historical data files from Gallup Canada and other polling companies, public-domain files such as the Canadian National Election Surveys and metadata records from the Inter-University Consortium for Political and Social Research (ICPSR).

Searching at the variable level (survey questions) is enabled across thousands of datasets in a growing number of collections. Researchers using <odesi> can perform basic tabulation and analysis online, and download most of the datasets into statistical software for further analysis. Datasets are delivered through the Scholars Portal platform which is already familiar to students and researchers at Ontario universities.

DataCite
www.datacite.org/
A group of research libraries and technical information providers, including CISTI, have established a partnership to improve access to research data on the internet. The goal of this initiative is to establish a

not-for-profit agency that enables organisations to register research datasets and assign persistent identifiers to them, so that research datasets can be handled as independent, citable, unique scientific objects.

As a first step, DataCite will promote the use of Digital Object Identifiers (DOI) for datasets. A DOI is used to cite and link to electronic resources (text as well as research data and other types of content). The DOI differs from other reference systems commonly used on the Internet, such as the URL, since it is permanently linked to the object itself, not just to the place in which the object is located. The long-term vision of the partnership is to support researchers by providing methods for them to locate, identify, and cite research datasets with confidence.

Archiving and Preservation

For research data to be re-used and/or re-purposed they must be integrated into an enduring institutional environment supported by a digital repository. The preservation of research data requires the active management of data over its entire lifecycle and involves activities such as "appraising, selecting, depositing or ingesting data into a repository, ensuring authenticity, managing the collection of data and metadata, refreshing digital media, and migrating data to new digital media."[7] The 2008 Gap Analysis found that research data in Canada are not being systematically preserved. It states, "Most of the data collected through research are not deposited into data repositories and few if any repositories have full preservation capacity as defined by TDR (Trusted Digital Repository) status."[8] Although data in certain disciplines are being collected by national agencies, this represents only a small minority of datasets created through research activities in Canada.

There has been growing interest within the library community in becoming more engaged in preservation of research data. Many data preservation activities are evolving out of existing institutional repository initiatives. However, there are limitations to the level of interoperability possible using current institutional repository software. Existing IR platforms do not yet have the functionality required for data to be tagged at the element level, something that is needed for interoperability and re-use of datasets. In addition, because research data are highly heterogeneous it is unlikely that any single repository could collect the range of data types created at any given single higher education institution. In *No Brief Candle*, Rick Luce, director of Emory University, recommends that initially, research libraries could best contribute by assisting with the preservation of smaller-scale data sets that arise from the work of local or domain-specific research groups.

Examples:

Collecting Data Using an Institutional Repository at MIT
http://pledge.mit.edu
The PLEDGE project at MIT is aimed at developing a mechanism to archive, preserve, and provide access to data using the DSpace repository software. Specifically, they are using Dspace MIT studies that are in the Institute for Quantitative Social Science (IQSS) Dataverse Network. The Libraries assist faculty with data preparation and provide assistance with data deposit. With the latter MIT researchers have three main options: DSpace, the Harvard-MIT Data Center (which provides the customized IQSS DVN), and the Inter-university Consortium for Political and Social Research (ICPSR) data archive. While Library staff promote all of these deposit options, they are also interested in providing a means of making faculty

[7] Research Data Strategy Working Group. pg. 3
[8] Research Data Strategy Working Group. pg. 7

produced research discoverable in all three systems thereby overcoming the dilemma of end-users having to search multiple places to find the data.

Network of Data Repositories at the eCrystals Federation Project
http://wiki.ecrystals.chem.soton.ac.uk/index.php/Main_Page
This project will establish a solid foundation of crystallography data repositories across an international group of 14 partner sites, with metadata harvested by a number of aggregator services. Partners will work together to: harmonise the metadata application profiles from repositories operating on different platforms (EPrints, DSpace, Fedora & ReciprocalNet); investigate aggregation issues arising from harvesting metadata from repositories within the networked information environments in other countries (EU, USA & Australia); enable the Federation of institutional archives to interoperate with international subject archives (IUCr and CCDC) and other third party harvesters; address areas of data curation, preservation and provenance by application to an operational Federation and through the provision of a testbed for the DRAMBORA Toolkit; provide a co-ordinated advocacy and training approach to repository deployment.

DISC-UK DataShare Project
http://www.disc-uk.org/datashare.html
DISC-UK DataShare project, led by EDINA and the Edinburgh University Data Library, with partners at the Universities of Southampton and Oxford, has advanced the current provision of repository services for accommodating datasets in the UK. By working together across four universities and internally with colleagues already engaged in managing open access repositories for e-prints, this partnership introduced and tested new models of data sharing and archiving at UK research institutions. By supporting academics within the four partner institutions who wish to share datasets on which written research outputs are based, this network of institution-based data repositories has developed a niche model for deposit of 'orphaned datasets' heretofore filled neither by centralised subject-domain data archives/centres/grids nor by e-print based institutional repositories (IRs).

Virtual Research Environments

E-research is described as a new research methodology, fueled by networked capabilities and vast amounts of data. E-research is fundamentally altering the ways in which researchers carry out their work, the tools they use, the types of problems they address, and the nature of the documentation and publication that results from their research. Virtual Research Environments (VREs) are part of the e-research infrastructure. They provide a framework of resources to support underlying research activities and are defined as "a set of online tools, systems and processes interoperating to facilitate or enhance the research process within and without institutional boundaries."[9]

Libraries' involvement with e-research and VREs is relatively nascent. As pointed out by Cliff Lynch, VREs can be an "uncomfortable fit with the rather passive tradition of libraries."[10] In addition, VREs are often cross-institutional and support collaborative research activities amongst geographically dispersed researchers, presenting a challenge for library participation. If libraries can overcome these barriers, they could be well placed to contribute to the development of VREs on campus.

Examples:

[9] Definition taken from the OSI e-Infrastructure Steering Group in the UK: www.nesc.ac.uk/documents/OSI/vrc.pdf
[10] Lynch, C. 2008. pg.1

myExperiment.org
http://myexperiment.org
The myExperiment Virtual Research Environment supports the sharing of research objects used by scientists, such as scientific workflows. For researchers it is both a social infrastructure that encourages sharing and a platform for conducting research, through familiar user interfaces. For developers it provides an open, extensible and participative environment. The software that powers myexperiment.org is downloadable so that users can run their own myExperiment instance.

The DataStaR Data Staging Repository at Cornell University
http://datastar.mannlib.cornell.edu/
DataStaR is designed as a data staging repository. It functions as a temporary repository for research datasets (in any stage of completion). Researchers submitting their data are provided with a workspace that allows for controlled sharing with selected colleagues (or with the public if the researchers wish), and which also provides remote storage and back up for the data. Once they meet requirements for deposit in an external disciplinary data archive or repository, the data can be passed from the staging repository [DataStaR] for publication. Researchers placing their data in the staging repository are also provided the option of submitting "publication-ready" data to Cornell University's IR, eCommons.

Islandora: Vitrual Research Environment at University of Prince Edward Island
http://islandora.ca/
Islandora is an open-source project under way at the Robertson Library. Islandora combines Drupal and Fedora software to create a robust digital asset management system that can be used where collaboration and digital data stewardship are required, for the short and long term. Islandora provides a key service through its VRE (Virtual Research Environment), a special website used by researchers at UPEI and elsewhere to steward research data. An increasing number of institutions around the world are using Islandora for their own repository systems. VRE (Virtual Research Environment) is a collection of customized Islandora sites used by researchers at UPEI and elsewhere to steward research data.

The Way Forward

Any solution for data stewardship in Canada must address issues across the entire lifecycle of research data, while also taking into account the great diversity of data sets across disciplines. Clearly, no single institution can tackle this problem individually. A comprehensive approach will require that all stakeholders- researchers, funding agencies, government departments, universities and libraries- work together to define roles and responsibilities and "pursue higher-level stewardship goals important to the entire community."[11] That being said, libraries can begin to lay the groundwork for a more coordinated and comprehensive national approach.

In a recent article describing data curation programs in the US, Tyler Walters, from the Georgia Institute of Technology describes the US situation. "The curation of scientific research data at U.S. universities is a story of enterprising individuals and of incremental progress… A small number of libraries and data centers who see the possibilities of becoming 'digital information management centers' have taken entrepreneurial steps to extend beyond their traditional digital assets and include managing scientific and scholarly research data."[12] Similarly in Canada, libraries currently interested in implementing research data services will likely have to do so without the support of institution-wide policies or targeted government funding programs. However, given the growing international trend towards data stewardship, these programs could be the seeds for more formal initiatives in the future.

Ultimately, the existing gaps in research data stewardship in Canada present an opportunity for libraries to create a larger presence for themselves in the evolving e-research infrastructure. This review found numerous options for libraries to become more involved in the area of research data. Some recommendations for libraries that are considering moving forward are included below.

Needs assessment

In order to develop effective services, libraries should have a good idea of what their researchers' needs are and of the nature of the data assets at the institution. This kind of information can be garnered through a needs analysis. In 2007, the Joint Information Systems Committee in the UK funded the development of a Data Audit Framework (DAF) to assist with data needs assessment. The framework provides organizations with the means to identify, locate, describe and assess how they are managing their research data assets. It combines a set of methods with an online tool to enable data auditors to gather this information. This is a practical tool that can be used or adapted by Canadian libraries to assess researchers' needs and develop data services based on these needs.

Developing expertise

While certain activities, like awareness and advocacy, may not necessitate library staff to acquire significant new knowledge, it is clear that libraries will have to invest in skills development for staff involved in more complex research data services. Core competencies for librarians in the area of data management include: migrating data to new formats; building ontologies, hierarchical structures and interactive thesauri; metadata production; data organization and preservation; and the development of access and discovery tools. Some level of domain expertise is also desirable since, ideally, data specialists will have an understanding of the needs of a specific discipline.

In the past, data management skills have often been nurtured in-house over time from local data scientists, or from collaborating with faculty on research projects. However, on the job training may not be sufficient for libraries aiming to implement robust research data services. There are a handful of accredited graduate-level academic programs and professional training and education programs available

[11] Research Data Strategy Working Group. pg. 7
[12] Walters, T. pg. 1

for librarians outside of Canada. For example, The Data Curation Education Program (DCEP), at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaigna prepares students to work in the area of data curation. They also host summer institutes aimed at training working librarians. There are a number of other training opportunities in the US and UK. It is also expected that the availability of these types of programs will increase given the growing demand. In Canada, the Canadian Association of Research Libraries (CARL), the Canadian Research Data Strategy Working Group, and Carleton University Library are sponsoring a number of research data workshops for librarians across Canada.

New organizational models

A 2008 report commissioned by JISC in the UK argues that support for data-intensive research will require a 'strategic repositioning' of the library.[13] Research libraries have traditionally been structured and staffed around disciplines. In contrast, e-Research embraces multidisciplinary approaches. According to Rick Luce, e-Research often requires virtual teams to form dynamically in the initial planning phases of a research project, work on a project, and then morph into something else when a less intense presence is needed. This entails fluid staffing structures and a more dynamic structural model than the current practice of assigning departmental or subject liaisons.[14] Luce envisions a future in which libraries have the capacity to create multi-skilled information management teams on the fly and embed librarians within research teams or departments.

Changes in research library structures ought to be driven by the needs of the research communities with whom they are working. It will require a shift in focus from delivering products to supporting processes. Because the data life cycle begins early in the scientific process, new library services should also be developed around the early research stages of a research project, as well as downstream services that address data preservation and reuse.[15]

Cost models and making a business case

The costs of most research data services are not insignificant. According to Rick Luce, "the widely decentralized and nonstandard mechanisms for generating data of every type and format imaginable" make assessing the costs of curation and preservation of data extremely complicated.[16] In 2008, JISC funded a study to investigate the costs of preserving research data. *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities* provides a methodological foundation for assessing the costs involved with collecting research data. The study shows that majority of the costs involved are staff (approximately 85%) and equipment (approximately 13%). It is recommended that libraries in the process of developing a cost model for data services consult this report.

Developing a business case for data services can be very challenging. Benefits cannot always be directly associated with improved research production within a particular research unit, but rather tend to be used by a research community in general. However, indirect benefits include raising the profile of an institution internationally. Other benefits, which are discussed in more detail in the companion document, *Research Data: Unseen Opportunities*, include maximizing investment in data collection; reducing duplication of data collection costs; increasing transparency of the scientific record; and enabling new scientific discoveries.

[13] Swan, A. pg. 1
[14] Luce, R. pg. 1
[15] Gold, A. pg. 1
[16] Luce, R. pg. 1

References

Gold, A. (2007). 'Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries'. *D-Lib Magazine 13, (9/10*). 2007. [Available at: www.dlib.org/dlib/september07/gold/09gold-pt2.html]

Harnessing the Power of Digital Data for Science and Society. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Sciences and Technology Council, January 2009.

Joint Task Force on Library Support for E-Science. *Agenda for Developing E-Science in Research Libraries. Association of Research Libraries*. Association of Research Libraries, November 2007. [Available at: http://www.arl.org/bm~doc/ARL_EScience_final.pdf]

Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. National Science Foundation, September 2005. [Available at: http://www.nsf.gov/pubs/2005/nsb0540/]

Luce, Richard E. 'A New Value Equation Challenge: The Emergence of eResearch and Roles for Research Libraries'. In *No Brief Candle: Reconceiving Research Libraries for the 21st Century*. CLIR Report 142, August 2008. [Available at: http://www.clir.org/pubs/reports/pub142/luce.html]

Lynch, C. The Institutional Challenges of Cyberinfrastructure and E-Research. *EDUCAUSE Review, 43(6*). 2008. [Available at: www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume43/TheInstitutionalChallengesofCy/163264]

Lynch, Clifford. "Where Do We Go From Here? The Next Decade for Digital Libraries". *D-Lib Magazine* 11(7/8), July/August 2005. [Available at: www.dlib.org/dlib/july05/lynch/07lynch.html]

Lyon, L. *Dealing with data: Roles, rights, responsibilities and relationships: Consultancy report.* UKOLN, University of Bath, UK. June 17, 2007. [Available at: www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf]

Martinez-Uribe, L. *Findings of the scoping study interviews and the Research Data Management Workshop: Scoping digital repository services for research sata management.* University of Oxford, UK. May 27, 2008. [Available at: www.ict.ox.ac.uk/odit/projects/digitalrepository/]

Neil Beagrie, Julia Chruszcz, and Brian Lavoie. Keeping Research Data Safe: A Cost Model and Guidance for UK Universities. Joint Information Systems Committee, 2008. [Available at: http://www.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx]

Research Data Strategy Working Group. *Stewardship of Research Data in Canada: Gap Analysis.* Canada Institute for Scientific and Technical Information, October 2008 [Available at: http://data-donnees.gc.ca/docs/GapAnalysis.pdf]

Shearer, Kathleen. *Research Data: Unseen Opportunities*. Canadian Association of Research Libraries, 2009 [Available at: www.carl-abrc.ca/about/working_groups/pdf/data_mgt_toolkit.pdf]

Swan, Alma and Sheridan Brown. *Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs*. Joint Information Systems Committee, July 2008. [Available at: http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx]

To Stand the Test of Time: Long-term Stewardship of Data Sets in Science and Engineering. A Report of the ARL/NSF Workshop on Long-Term Stewardship of Digital Data Collections, September 26-27, 2006. Association of Research Libraries [Available at: www.arl.org/pp/access/nsfworkshop.shtml]

Wilbanks, John. "I Have Seen the Paradigm Shift, and It Is Us". *The Fourth Paradigm: Data Intensive Scientific Discovery*. Microsoft Research, 2009. pg. 209-214. [Available at: http://research.microsoft.com/en-us/collaboration/fourthparadigm/]

Walters, Tyler O. "Data Curation Program Development in U.S. Universities: The Georgia Institute of Technology Example". *International Journal of Digital Curation*, 4 (3), 2009. [Available at: http://www.ijdc.net/index.php/ijdc/article/view/136]

Wusteman, Judith.Virtual Research Environments: What Is the Librarian's Role? *Journal of Librarianship and Information Science*, 40 (2) June, 2008.