National Research Data Initiatives in Canada

Lynn Copeland
Former Dean of Library Services
Simon Fraser University and Project Co-coordinator CNCDI
Vancouver, BC, Canada
copeland@sfu.ca

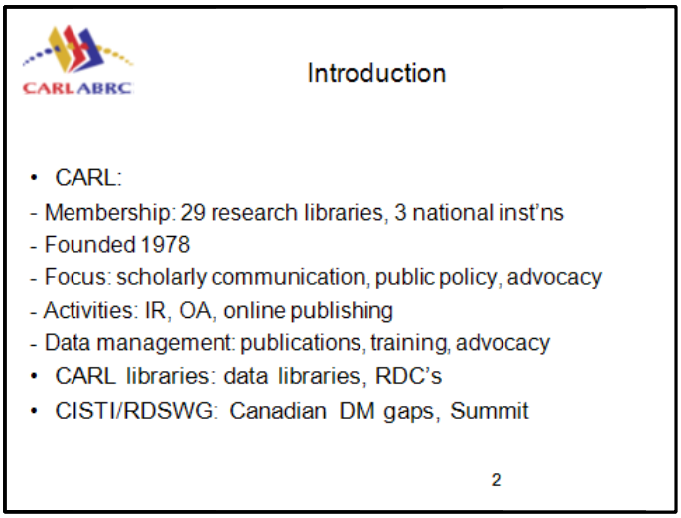
Brent Roe
Executive Director
Canadian Association of Research Libraries
Ottawa, ON, Canada
brent.roe@carl-abrc.ca

Session: 116 — *The role of libraries in data curation, access and preservation: an international perspective* — Science and Technology Libraries

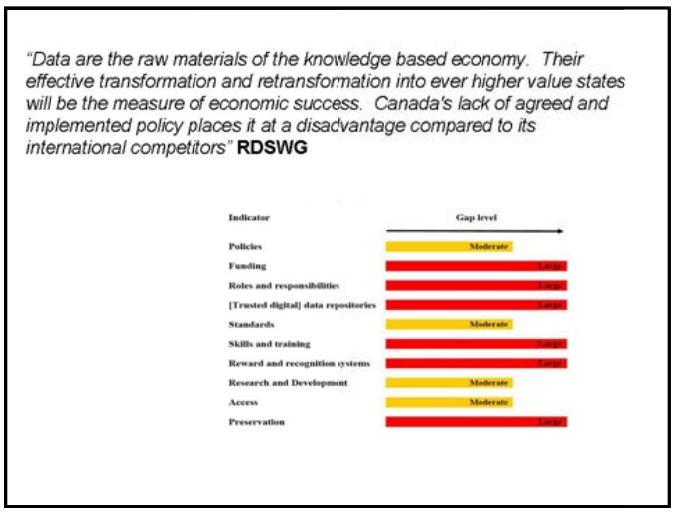
Abstract :

Canadian Association of Research Libraries (CARL) member libraries have played an active role in providing research data curation and access on campus, and in various regional and national research data initiatives. CARL has been active as an organization, producing information pamphlets and training. The Canadian Institute for Scientific and Technical Information (CISTI) has also been proactive in a number of issues related to national research data preservation and curation. In 2010-11, CARL investigated the potential for a nationally-funded research data preservation initiative. Wide consultation and discussions were initiated. A focus group of researchers from a broad range of disciplines and consultations with funding agencies, national IT groups, campus VPs Research and others were successful in raising the profile of CARL's individual-member library and collective activities and promoting the importance of a Canadian national research data curation and preservation strategy and service. Ultimately, owing to the straitened funding environment, no proposal was submitted. Nevertheless, the initiative was successful in engaging key players (IT organizations, libraries, government, universities, Library and Archives Canada, CISTI) and raising libraries' profiles in the key role of research data access, curation and preservation in partnership with IT organizations. Subsequent to the discontinuance of the initiative, CARL continues to be involved in activities on the national front. A national IT summit will be held in June 2012. The CISTI RDSWG is in the process of forming Research Data Canada to promote and fundraise for a national research data infrastructure. Other initiatives by CARL, CISTI, national IT organizations and individual and regional libraries continue.

Introduction



Formed in 1976, the Canadian Association of Research Libraries includes 29 university research libraries and three national research library organizations. CARL is engaged with the continuing transformation of scholarly communication, advocacy for favourable outcomes in the federal public policy environment, and strengthening and promoting Canada's research libraries [1]. CARL has support initiatives related to institutional repositories, open access, and online publishing. In particular, CARL has for some years focused on data management, producing several reports and pamphlets [2] and supporting an education program and e-science institute for research. CARL has also advocated in its submissions to federal budget consultations and elsewhere the importance of national support for research data archiving. Many individual Canadian research libraries have provided data services through a Research Data Library as far back as the early 1970s, focusing on purchased or licensed Social Sciences, Research Data Centre, and financial data and services, in many instances through collaborative development and support of storage and retrieval tools.



The Canadian Institute for Scientific and Technical Information (CISTI) [3], the science library arm of Canada's National Research Council, has also played a significant role in the advancement of research data stewardship through its Research Data Strategy Working Group (RDSWG) under the leadership of Pam Bjornson, Director General of CISTI, with representatives from a range of organizations with a stake in research data management [4]. The RDSWG wanted to consider how to make progress on some of the recommendations from the National Consultation on Access to Scientific Research Data (NCASRD)[5] of 2004 and subsequent consultations that had only limited success. The RDSWG identified significant gaps in Canada's research data infrastructure (Table 1), noting that "data are the raw materials of the knowledge-based economy. Their effective transformation and retransformation into ever higher value states will be the measure of economic success. Canada's lack of agreed and implemented policy places it a disadvantage compared to its international competitors."[6] Moderate gaps were identified in policies, standards, research and development and access; large gaps were identified in funding, roles and responsibilities, trusted digital repositories, skills and training, reward and recognition systems and preservation. In no area were the gaps small or nonexistent nor did Canada excel in any area.

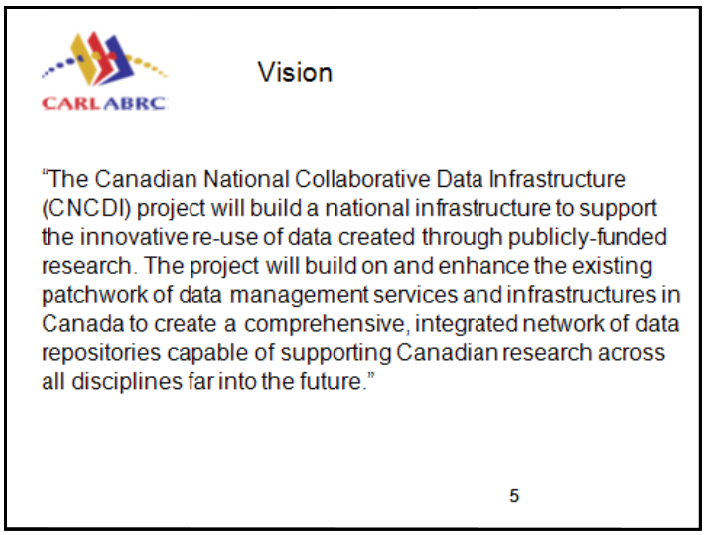
The Canadian National Collaborative Data Infrastructure Initiative

In 2010, there was a call for proposals from the Canada Foundation for Innovation (CFI) [7], which had previously funded the national Research Data Centres, the Canadian Research Knowledge Network (CRKN; the major Canadian academic digital content licensing consortium) electronic collections and Synergies e-journals and other related initiatives based in Canadian research libraries. A number of data management projects were under way or funded at the individual institutional or cross-institutional level. These included the International Polar Year Data Centre Assembly Network project, involving the University of Alberta, the University of Waterloo and Scholars Portal (the digital licensing and service organization of the Ontario Council of University Libraries) as well as three federal government departments; a project funded at Simon Fraser University (SFU) to preserve research data in several social science areas; a CFI proposal involving SFU and the University of Montreal building on the Synergies project, as well as others.

With this background, CARL began to investigate the possibility of a national research data initiative and submission in the form of an initiative it called the “Canadian National Collaborative Data Infrastructure” (CNCDI). Ultimately, CFI decided not to fund national research infrastructure projects through the funding programs active at that time, but meanwhile CARL had achieved some success in building connections and mutual understanding with related national IT organizations, granting councils, researchers, and university administrations. These efforts have enabled further initiatives which will, it is hoped, lead ultimately to financial support at a national level.

The CNCDI vision still pertains:

> “The benefits of sharing and repurposing research data can be powerful. They range from the mapping of the human genome, to the identification of new planets and solar systems, to the discovery of oil deposits through the re-examination of satellite data. Imagine how many more new discoveries could be made if all research data were shared in this way! In the words of a European Union high-level expert group, “We are on the verge of a great new leap in scientific capability, fuelled by data” [Riding the wave, pg.9]. Indeed, the availability of digital data is dramatically redefining the nature and scope of the research endeavor across all domains in the 21st century.
>
> We envision a future in which Canada capitalizes on the trend towards data intensive research: a future in which Canada is a world leader in research and technology, and a locus for entrepreneurial innovation and creativity. This future is

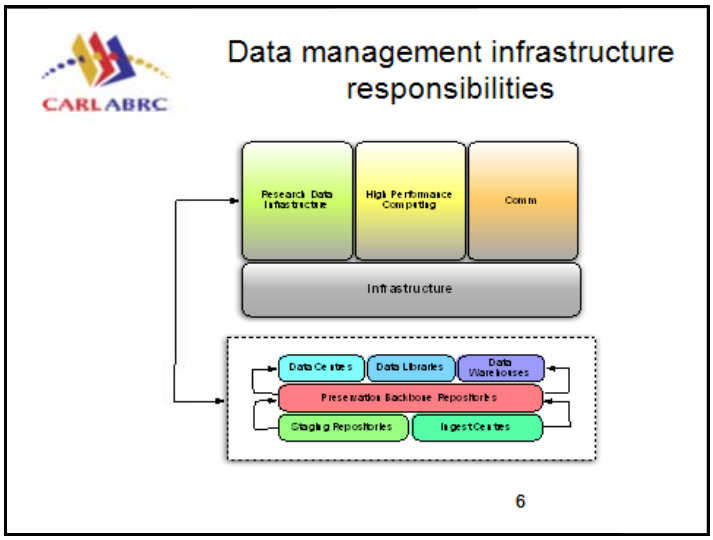


achievable, but it will require the development of a research data management infrastructure on a national scale. The infrastructure must be flexible and reliable; secure, yet open; local and global; and affordable, yet high-performance. With this infrastructure, research data, with appropriate security and privacy controls, will become widely available to all sectors - academia, industry, government, and the public at large - facilitating a cross-fertilization of ideas and disciplines that will produce novel solutions, support the development of new products, and promote greater understanding of complex problems. With such an infrastructure, Canada will be poised to become a major player in the global research arena."[8, p. 3]

Consultations

CARL engaged in a series of consultations with librarians, granting councils and CFI, with national IT groups (specifically CANARIE, Canada's national high-speed digital research network [9]; CUCCIO, the Canadian University Council of Chief Information Officers [10]; and Compute Canada, Canada's national high-performance computing consortium [11]) that support university research and with key individuals in government and on campus. The results of these consultations formed the basis of CNCDI's early conception.

CARL articulated a model for the services and facilities of a research data management infrastructure, including the



responsibilities that library and IT partners might respectively assume (Figures 2 and 3). Three tiers of Research Data Infrastructure responsibility were identified and a draft business plan including rough costs for each tier was

developed. The tiers were: redundant storage repositories; disciplinary ingest centres; and microservice sites. The three tiers identified were conceptual, with the expectation that an institution might assume responsibility for more than one tier and that the specifics for each tier would evolve in the initial stages of the project [12].

A focus group of researchers spanning a variety of disciplines, lengths of career experience, regions, languages, and genders was held in March 2011 and a report subsequently produced [13]. The conclusions from the focus group report were and remain important in the framing of a Canadian data curation, management and preservation infrastructure.

Issues identified by the focus group revolved around privacy, ethics and accessibility; data culture and advocacy; data rights, ownership and intellectual property; sustained resourcing and support; metadata and standards (quality and

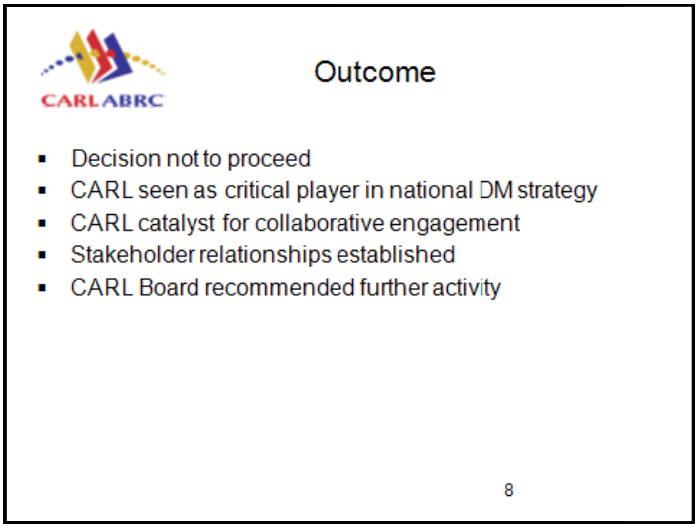
interoperability); and, in terms of the project itself, the need for coordination and definition of project scope and the technical challenges involved.

The group identified possible negative responses to the project, relating to fears of 'big brother' and privacy; a belief that data have little value beyond the original research project; a lack of understanding of libraries' roles; a concern about the negative impact on research of the commodification of data; and a belief that the commercial sector could do it more quickly.

Potential impacts of *not* preserving research data included the loss of control over valuable research data needed for further research; the threat that universities might be unable to fulfill their mission of supporting leading edge research; the concern that there will be a lack of sufficient data to develop evidence-based practices and policies; the threat that jurisdictional issues of ownership and lack of interoperability will confound data access; the negative impact on society, economy and population health (and other negative downstream effects); loss of competitiveness for Canada internationally in the research arena; data not being used to their full potential; and, even if preserved, if done independently, the worry that data sets would not be interoperable.

The positive impacts on research of preservation and curation of data included greater access to research data for research and policy makers; increased speed of discovery; stronger links between research and policy/practice; the expectation that innovative tools and technologies would be developed to make use of data in new ways; the fostering of greater collaboration between industry and academic research. Discipline-specific data management resources and supports will be readily available. There will be greater data sharing, greater use of standards and data integration, incentives for researchers to share their data, acknowledgement when data are used by others, protocols for managing rights and ownership, general understanding that data are an important public asset, protocols to ensure that data are protected and anonymized, and consistent policies and practices across national funding agencies. Researchers will spend time on research rather than on the "mechanics" of managing data. There will be a multidimensional complete classification scheme. There will be endless possibilities for new research.

Discontinuance of CNCDI



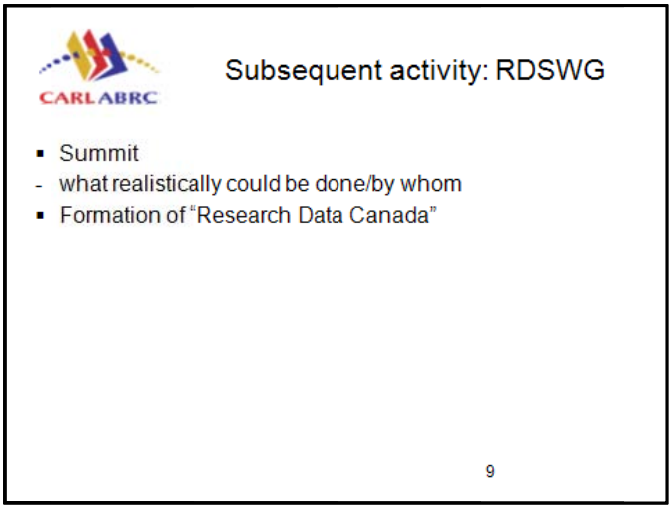
Ultimately, with no national project funding in the current CFI round and increased focus on specific research projects, the decision was taken not to proceed with an application for CNCDI at this point. Nevertheless, the need that led to the inception of the CNCDI project had not diminished in importance. With CARL now seen as a critical player in developing a national strategy for research data infrastructure – one which must include the three pillars of high speed communications networks, high performance computing and data management and preservation services -- work that CARL had done relating to the latter pillar laid the groundwork for demonstration projects and for future funding opportunities.

Over the course of the CNCDI project, CARL was a catalyst for collaborative engagement in the issue of research data management at the national level. Substantial strides were made in establishing relationships with other stakeholders, raising the profile of the issue and articulating the data management and preservation role that CARL libraries can and should play in the research data arena. In January 2012, the CARL Board accepted the Final Report of the project, recognizing the significant progress of the CNCDI initiative and the continuing imperative to to work with the other stakeholder groups to build on a heightened awareness and understanding of the need to steward Canada's research data [13].

New Developments in the Landscape of Canadian Research Data Management

Upon reviewing the CNCDI project Final Report and with input from its authors, the CARL Board proposed a set of actions to guide CARL research data management initiatives in 2012 and beyond [15], believing that collaboration is one of the hallmarks of Canadian research libraries and that CARL's experience in this regard, and its member libraries' expertise in data management and preservation, would help to facilitate forward movement in the development of a national research data infrastructure on other fronts than CFI funding. Individual and regional research library data initiatives continue and CARL continues to be actively engaged in a number of initiatives discussed below.

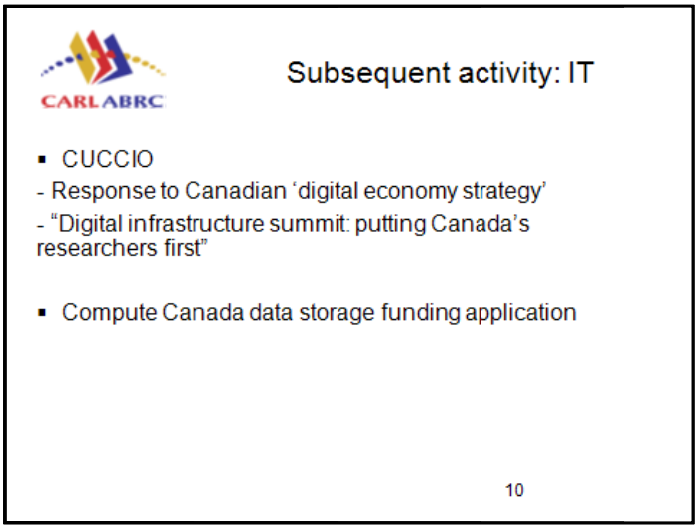
The Research Data Strategy Working Group and the Canadian Research Data Summit



Shortly before the CNCDI initiative wound down, the CISTI RDSWG (of which CARL is a member) hosted a "Canadian Research Data Summit," September 14-15, 2011 [16]. Rather than focusing on clarifying the "data issue," the Summit was intended to identify, in consultation with all community stakeholder groups, what realistically could and should be done and by whom. The Summit was successful in that a series of well received recommendations were produced and there was a sense of willingness by many participant groups to support the effort tangibly [17].

In the immediate wake of the Summit, the RDSWG decided that it would be both necessary and opportune to seek support from the stakeholder organizations in order to form a higher-profile group that could act as a coordinating group and raise funds to sponsor projects important to furthering the ability of Canadian researchers to archive their data so that it could be re-used by others. It is envisioned that the new organization will be called "Research Data Canada" and will, it is anticipated, obtain a portion of its initial funding from such groups as the federal granting councils, the associations of research libraries and university CIO's, and the national high-speed research network and high-performance computing agencies [18].

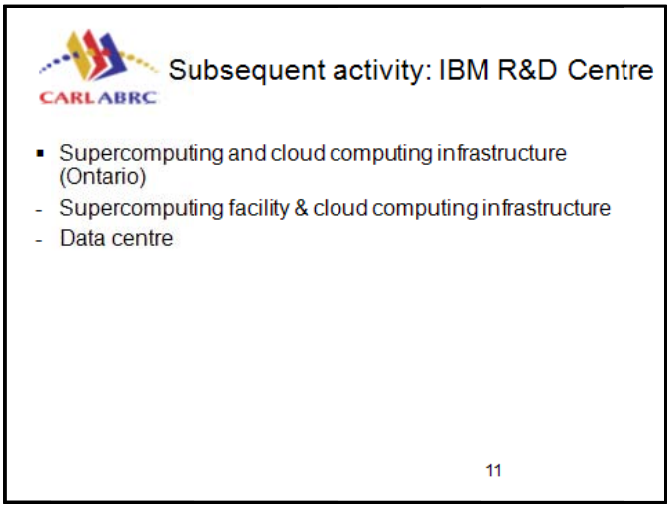
The CUCCIO Summit: Digital Infrastructure for Research



In 2010, the university CIO's (as CUCCIO), collaborated with several other university research IT support organizations in responding to a national public consultation on a Canadian "digital economy strategy"[19]. Rather

than the existing piecemeal approach to Canadian research infrastructure, they felt that what was needed was a more integrated national approach to a "research, education and innovation (REI) infrastructure," a part of this infrastructure being a repository system for research data [20]. When it became clear in fall 2011 that major funding would not be found in the near term for the CNCDI initiative, CUCCIO began to plan for a summit meeting, "Digital Infrastructure Summit 2102: Putting Canada's Researchers First," to be held June 13-14, 2012 [21]. Focused on the broader concept of what was now called a "digital infrastructure for research," which does include research data management, the Summit and the new organization that may be engendered by it will actually have a very significant data management aspect. It is not yet clear how this potential new group will find a place with the emerging Research Data Canada, but with the same major stakeholder organizations in both of these forums, these parallel activities are indicative of a strongly felt need for an organization that can concentrate on these issues.

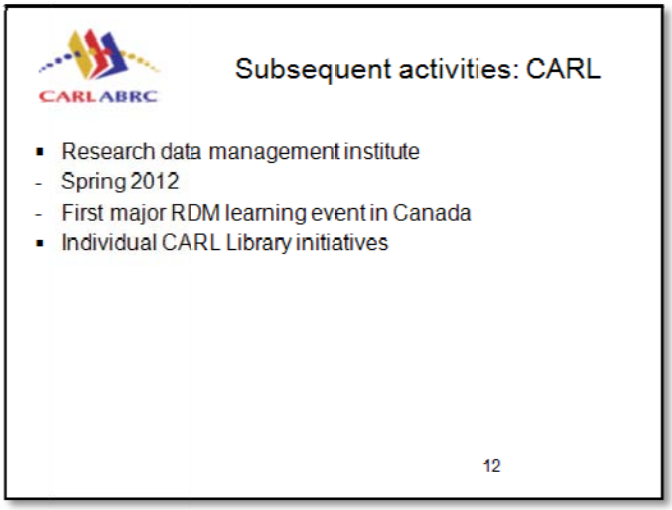
The IBM Canada Research and Development Centre



On April 10, 2012, the creation of a new supercomputing and cloud-computing infrastructure was announced, with $175 million from IBM, $20 million from the federal government and $15 million from the Ontario provincial government [22]. Seven Ontario universities will share a supercomputing facility at the university of Toronto, a cloud computing hub at Western University (London, Ontario), and a "data centre" to be built in Barrie, Ontario, by fall 2012. According to the IBM "Backgrounder" on the project, "these investments will help IBM and university researchers use modern high performance and cloud computing capabilities to better manage and apply "big data" to solve critical challenges in our cities, healthcare, water and energy management systems"[23].

The CARL Research Data Management Institute Being Planned for 2013

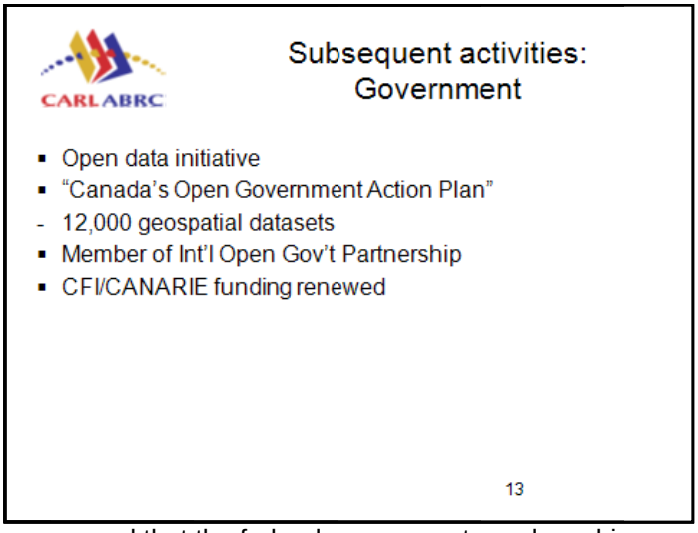
Recognizing that the most important contributions of the libraries to data management will be working with



researchers to help them structure and describe the data they generate and to locate and use the data that is archived for re-use, CARL is developing a Research Data Management Institute for academic librarians across Canada so as to increase librarians' competencies in data management [24]. It is expected that the first Institute will be in Spring 2013. Other than an introductory one-day general workshop on data held in 2009 and 2010, also developed by CARL, the Institute will be the first major research data management learning event ever run in Canada.

The Open Data Initiative of the Federal Government

During 2010, a small group of Government of Canada staff members developed a simple web portal to give easier public access to federal government departments' datasets that were already, at least in theory, publicly available. This portal gained some favourable attention internally and, in March 2011, Treasury Board President Stockwell Day



announced that the federal government was launching a one-year pilot project in Open Data whereby the government would make some 260,000 datasets available to the public through the "Open Data Portal" as an element of the government's "digital economy strategy" and its commitment to Open Government [25]. The vast majority of the data sets made available were geospatial in nature, but the government promised to add more over time in other areas.

In March 2012, the new Treasury Board President Tony Clement, in announcing "Canada's Open Government Action Plan" extended the Open Data Portal project for another three years and announced the addition of additional datasets [26]. At time of writing, there are almost 12,000 non-geospatial datasets available for searching in the Portal. Somewhat ironically, in 2010 the same Mr. Clement had also cancelled the mandatory long-form national census, replacing it with a voluntary survey [27]. Nevertheless, the federal government has facilitated public access to government datasets and seems to understand the value of that both for university research and for private commerce and industry.

The Data Storage Funding Application of Compute Canada

In the fall of 2011, Compute Canada, the national high-performance computing organization, submitted an application to CFI under its "Leading Edge Fund," in part for substantial increased data storage capacity [28]. CFI will not be announcing the successful proposals until fall 2012.

Renewed Government Funding for CFI and CANARIE

While the 2012 federal budget is an austerity budget, with an expected loss of 19,000 federal civil service jobs, the government nevertheless did maintain its support for science and research [29]. Among other provisions, CFI received a new $500 million to be spent over five years beginning in 2014-2015, which may yet allow the creation of a funding program that could propel the development of a national data management infrastructure. As well, CANARIE, the national high-speed digital research network, received a two-year $40 million funding commitment, which is important, as they are a key component of any national data management infrastructure [30].

Conclusion

CISTI, Canadian research libraries and national IT groups have worked together over a number of years toward a national research data infrastructure. The CNCDI was an important part of this process and CARL continues to play an active role in research data infrastructure in Canada, working toward opportunities for national funding in, it is hoped, the near future.

References

1. *Home*. Canadian Association of Research Libraries, 2011. Web site. April 23, 2012. <http://www.carl-abrc.ca>.
2. Available at http://carl-abrc.ca/en/scholarly-communications/data-management-sub-committee.html#papers and http://carl-abrc.ca/fr/communication-savante/gestion-des-donnees-sous-comite.html#documents.
3. Canada Institute for Scientific and Technical Information. *NRC National Science Library*. National Research Council Canada, March 26, 2010. Web site. April 23, 2012. <http://cisti-icist.nrc-cnrc.gc.ca/eng/ibp/cisti/index.html>.
4. Research Data Strategy Working Group. *Research Data Strategy*. Canada Institute for Scientific and Technical Information, January 30, 2012. Web site. April 23, 2012. <http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/>.
5. Research Data Strategy Working Group. *National Consultation on Access to Scientific Research Data (NCASRD)*. Canada Institute for Scientific and Technical Information, March 8, 2012. Web site. April 23, 2012. <http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/ncasrd/index.html>.
6. Research Data Strategy Working Group. *Stewardship of Research Data in Canada: A Gap Analysis*. Canada Institute for Scientific and Technical Information, October 2008. PDF Web document. April 23, 2012. <http://publications.gc.ca/collections/collection_2009/cnrc-nrc/NR16-123-2008E.pdf>.
7. *INNOVATION.CA.* Canada Foundation for Innovation, 2012. Web site. April 23, 2012. <http://www.innovation.ca/en>.
8. Copeland, Lynn and Martha Whitehead. *Canadian National Collaborative Data Infrastructure Project: Final Report.* Canadian Association of Research Libraries, January 2012. PDF Web document. April 23, 2012. < http://carl-abrc.ca/uploads/pdfs/carl_cncdi_final_report.pdf>.
9. *Canada's Advanced Research and Innovation Network*. CANARIE, 2009. Web site. April 23, 2012. <http://www.canarie.ca>.
10. *CUCCIO-CDPIUC*. Canadian University Council of Chief Information Officers, October 20, 2011. Web site. April 23, 2012. <http://www.cuccio-cdpiuc.ca/en/index.php>.
11. *Home*. Compute/Calcul Canada, 2011. Web site. April 23. 2012. <https://computecanada.ca/index.php/en/>.
12. *Canadian National Collaborative Data Infrastructure Project Preliminary Business Plan – Draft 3.* Canadian Association of Research Libraries, September 16, 2011. PDF Web document. April 23, 2012. < http://carl-abrc.ca/uploads/pdfs/CNCDI_preliminary_BusinessPlan-draft_3.pdf>.
13. Shearer, Kathleen. Canadian National Collaborative Data Infrastructure Project: Report of Research Data Symposium: March 10 & 11, 2011. Canadian Association of Research Libraries, April 15, 2011. PDF Web document. April 23, 2012. < http://carl-abrc.ca/uploads/pdfs/researcher_data_consultation_march_11-final.pdf>.
14. Copeland, Lynn and Martha Whitehead. *Canadian National Collaborative Data Infrastructure Project: Final Report* (see above).
15. Board of Directors. *CARL Research Data Management Initiatives: Next Steps*. Canadian Association of Research Libraries, January 2012. PDF Web document. April 23, 2012. <http://carl-abrc.ca/uploads/pdfs/cncdi_next_steps.pdf>.
16. Research Data Strategy Working Group. *The 2011 Canadian Research Summit*. Canada Institute for Scientific and Technical Information, January 30, 2012. Web site. April 23. 2012. <http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/events/data_summit_2011/index.html>.
17. Research Data Strategy Working Group. *Successful Canadian Research Data Summit held September 14-15*. Canada Institute for Scientific and Technical Information, September 20, 2011. Web site. April 23, 2012. <http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/news/data_summit.html>.
18. Research Data Strategy Working Group. *2011 Canadian Research Data Summit Final Report Released*. Canada Institute for Scientific and Technical Information, January 30, 2012. Web site. April 23, 2012. <http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/news/data_summit_report.html>.
19. Industry Canada. *Government of Canada Launches National Consultations on a Digital Economy Strategy*. Government of Canada, September 1, 2010. Web site. April 23, 2012. <http://www.ic.gc.ca/eic/site/ic1.nsf/eng/05531.html>.
20. CANARIE et al. "Canada's Digital Environment for Research, Innovation and Education," July 8, 2010. *The Digital Economy in Canada*. Government of Canada, March 30, 2012. Web document. April 23, 2012. <http://www.digitaleconomy.gc.ca/eic/site/028.nsf/eng/00437.html>.
21. Unpublished correspondence.
22. "Groundbreaking research collaboration creates a new economic cornerstone for Canada," April 10, 2012. *News*. IBM (Canada), n. d. Web document. April 23, 2012. <http://www.ibm.com/news/ca/en/2012/04/10/t770509v27527r12.html>.
23. Bendzsa, Carrie. "BACKGROUNDER—IBM Canada Research and Development Centre." *Ideas from IBM*, n. d. Microsoft Word document on Web. April 23, 2012. <http://www.ibm.com/ibm/ideasfromibm/ca/en/files/en_ca_research_network_backgrounder.doc>.
24. Unpublished CARL Data Management Subcommittee minutes.
25. Treasury Board of Canada Secretariat. "Minister Day Launches Open Data Portal," March 17, 2011. *Media*. Government of Canada, March 17, 2011. Web site. April 23, 2012. <http://www.tbs-sct.gc.ca/media/nr-cp/2011/0317a-eng.asp>.

26. Treasury Board of Canada Secretariat. “Minister Clement Releases Open Government Action Plan,” April 12, 2012. *Media*. Government of Canada, April 12, 2012. Web site. April 23, 2012. <http://www.tbs-sct.gc.ca/media/nr-cp/2012/0412-eng.asp>.
27. “Clement won’t back down on census,” July 22, 2010. CBC, 2012. Web site. April 23, 2012. <http://www.cbc.ca/news/politics/story/2010/07/22/statscan-census-tories-.html>.
28. Unpublished correspondence.
29. “Chapter 3.1: Supporting Entrepreneurs, Innovators and World-Class Research.” *Budget 2012*. Government of Canada, March 29, 2012. Web site. April 23, 2012. <http://www.budget.gc.ca/2012/plan/chap3-1-eng.html>.
30. “CANARIE’s Response: Kathryn Anthonisen,” April 9, 2012. *Home*. Canadian Science Policy Centre, n. d. Web site. April 23, 2012. <http://sciencepolicy.ca/news/canarie%E2%80%99s-response-kathryn-anthonisen>.

“Data are the raw materials of the knowledge based economy. Their effective transformation and retransformation into ever higher value states will be the measure of economic success. Canada's lack of agreed and implemented policy places it at a disadvantage compared to its international competitors” **RDSWG**

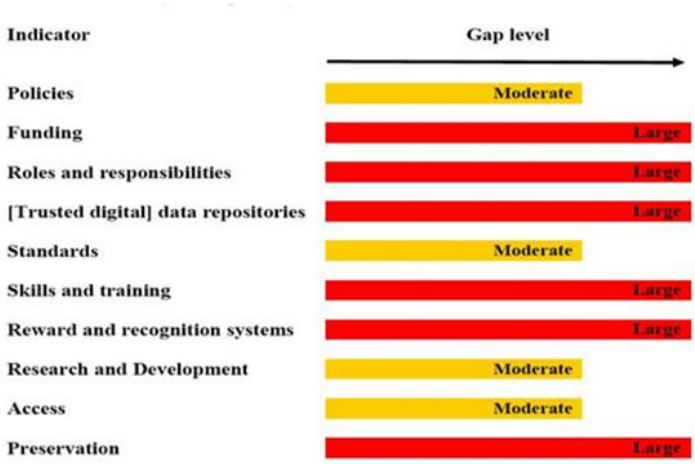


Table 1. CISTI RDSWG. Stewardship of Research Data in Canada: a Gap analysis

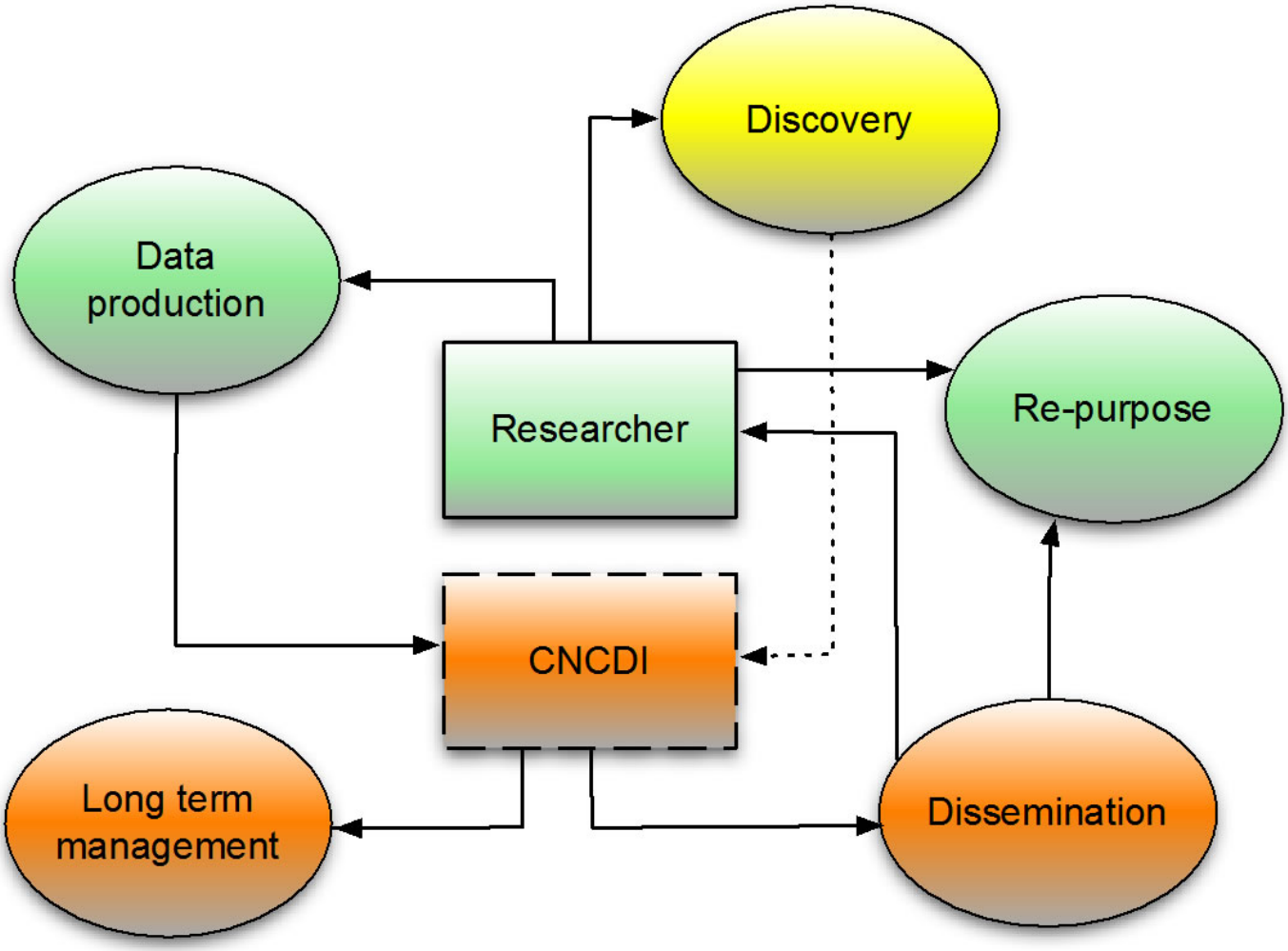


Figure 2. Data management infrastructure model

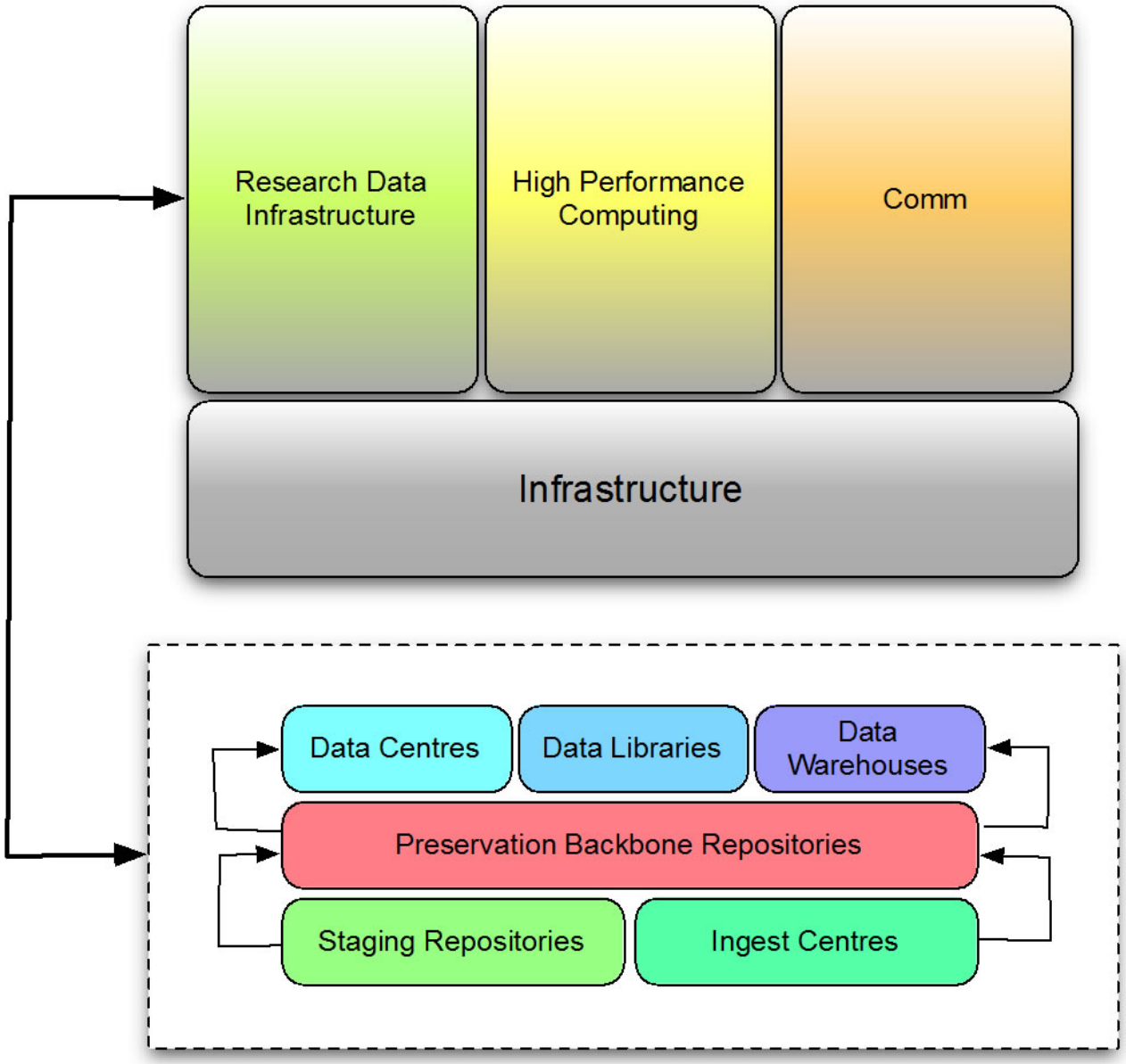


Figure 3. Data management infrastructure responsibilities