Research Data: Unseen Opportunities

In the last five years, the world produced more research data than has been created in all of human history.[i] Most of this data is in digital format, making it possible for researchers to re-purpose and use them in innovative ways not envisioned by those who first created them.

Research data are the factual records used as primary sources for research, and that are commonly accepted in the research community as necessary to validate research findings.

Right now in Canada the vast majority of research data is being lost because it is not being systematically managed. While certain disciplines and research projects have institutional, national, or international support for data management, this support is available for a minority of researchers only.

In 2004, 34 countries, including Canada, signed the OECD "Declaration on Access to Research Data from Public Funding".[ii] The premise of the declaration is that publicly funded research data should be openly available to the maximum extent possible. Following on this, a number of countries have begun to investigate how they can more systematically exploit the data created through research. Canada, however, has not yet responded with concrete actions to give its commitment to the OECD Declaration substance.

A coordinated and national approach to managing research data in Canada is required in order to derive greater and longer term benefits, both socially and economically, from the extensive public investments that are made in research.

Both the US and the UK governments have recently launched open data initiatives that increase public access to high value, machine readable datasets generated by their federal governments. These initiatives aim to expand creative use of government-generated data beyond the walls of government by encouraging innovative ideas, tools and web applications. It is expected that these initiatives will have large economic benefits, as businesses and individuals build new applications and services to analyze previously locked-up government data in new ways. In the UK, for example, they are anticipating an added value of £ 6 billion (GBP) from the initiative.[iii]

Data are the raw materials of a knowledge economy. Their effective transformation and retransformation into ever-higher value states is a major indicator of a nations economic success.

In 2009, the Australian government established the Australian National Data Service (ANDS) in order to improve the management of research data in that country. The aims of the ANDS Service are to influence national policy; inform best practice; and transform the disparate collections of research data around Australia into a cohesive collection of research resources.

Many organizations find out the hard way that the costs of maintaining data properly are insignificant compared to the expenses incurred in trying to replace or recreate lost data. For example, data from the Canada Land Inventory, a comprehensive multi-disciplinary land inventory of rural Canada, was very nearly lost due to neglect. It took four federal government departments over 5 years and many thousands of dollars to restore the data. It would have been nearly impossible to recreate this valuable data which has since sparked a number of new initiatives.

Since the 1960s, geographic information systems (GIS) has been transforming decision-making in industry, governments, and universities by bringing digital spatial data sets and geographic analysis to desktop computers. GIS is a technological field that incorporates geographical features with tabular data in order to map, analyze, and assess real-world problems, such as urban planning, asset management, and resource exploration. GIS has evolved into a huge industry worth billions of dollars.

Major Benefits of Data Management, Sharing, and Reuse

Accelerates scientific progress
Data sharing allows researchers to access and understand others' data and re-use them for their own scientific purposes, thereby speeding up the rate of new discoveries.

Avoids duplication of research
When a dataset is publicly available it is much less likely to be recreated, avoiding expensive and needless data collection/ production activities.

Ensures compliance with funding agency policies
A growing number of funding agencies demand that researchers and host institutions retain, manage and share their data upon completion of a research project.

Increases the visibility and impact of research
Data made visible through a data repository can dramatically increase the impact of that research.

Sharing research data has been associated with increased citation rates. For example, a study of citation rates for cancer clinical trials publications found that clinical trials that shared their data were cited about 70% more frequently than clinical trials that did not.[iv]

Enables replication and verification of research results
When data are archived and shared, results are repeatable and data can be used for re-analysis, backing up original research findings.

Enhances collaboration
Publicly available data enable researchers to better collaborate with each other by sharing data sets, research environments and tools.

Current Context

Approaches to the management of research data vary significantly according to discipline. Some fields, such as genomics, proteomics, high-energy physics, and astronomy have long-standing traditions of data archiving and sharing. Others, such as chemistry and the humanities and social sciences have less established traditions.

The International Polar Year (IPY) was the largest-ever international program of scientific research focused on the Arctic and Antarctic regions. It involved thousands of researchers from over 60 nations and produced large and diverse amounts of research data. IPY researchers are required to share their data with others in a timely manner and effectively manage their research data to ensure the data are accessible for future research. The lack of available data repositories is posing a challenge for many IPY researchers who are trying to fulfill these requirements.

Pioneering archives such as GenBank have demonstrated just how powerful legacy data sets can be for generating new discoveries; especially when data are combined from many laboratories and analyzed in ways that the original researchers could not have anticipated.

The Victoria Experimental Network Under the Sea (VENUS)[v], led by the University of Victoria, is a cabled ocean observatory. VENUS delivers real time data from the seafloor via fibre optic cables connected to instruments at the University of Victoria, BC, where they are archived. VENUS instruments have sent over half a trillion measurements back to the University of Victoria and, through the internet, to scientists and the rest of the world. To handle the huge volume of data, the University of Victoria has developed an advanced data management and archiving system that will read, store and make the data web accessible for at least the next 25 years to everyone-both the public and scientific communities- in order to advance ocean research in Canada.

In 2007, The Journal of Applied Developmental Psychology published a special issue to highlight how the Study of Early Child Care and Youth Development data sets had been used by researchers to address a range of research questions not envisioned in the original study plan. The data sets contain longitudinal data about the family, day care and school environments of over 1000 children, tracked from birth through age fifteen. Most of the publications by original study investigators focused on child-care issues. However, editors found that the original data have been used by hundreds of others investigating a wide range of other research problems, such as child development, women's employment patterns, quality of academic instruction, family social and psychological processes, child–adult interaction, and peer interaction.

End Notes

i. According to Tony Hey, former Director of the e-Science Core programme in the UK at Digital Preservation Coalition Forum. British Library Conference Centre, London UK, Wednesday 23 June 2004.
ii. From: http://www.oecd.org/document/0,2340, en_2649_34487_25998799_1_1_1_1,00.html
iii. From: http://www.conservatives.com/Policy/Where_we_stand/Technology.aspx
iv. Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308
v. From: http://www.venus.uvic.ca
vi. The full Gap Analysis is available at: http://data-donnees.gc.ca/eng/about/achievements.html
vii. Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers
viii. Heather A Piwowar, et. al. PLoS Med. 2008 September; 5(9): e183.Charles Humphrey (2004). Preserving Research Data: a time for action. Preservation of Electronic Records: New Knowledge and Decision-making. Ottawa: The Canadian Conservation Institute, pp. 83-90.

This brochure was created in June 2010 by Kathleen Shearer on behalf of the Canadian Association of Research Libraries' Data Management Subcommittee.

Canadian Association of Research Libraries
Morisset Hall
65 University Street Suite 239
Ottawa Ontario Canada
K1N 9A5
Phone: 613.562.5385
Facsimile: 613.562.5297
Email: carladm@uottawa.ca
www.carl-abrc.ca