OPEN SCIENCE: PAST, PRESENT AND FUTURE

Andrea Miller-Nesbitt
McGill Library Bibliothèque

OPEN SCIENCE OVERVIEW

According to prominent open science advocate Michael Nielsen, open science is based on the idea that "scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process"[1]. Sharing the research process has never been easier, as most data are born digital and stored electronically. Using a cyberinfrastructure of integrated science tools scientists are now able to conduct the research process openly from beginning to end.

There are many benefits to doing research openly, including accelerating the discovery process by making use of collective intelligence and citizen science, reducing the duplication of work thereby saving time and resources, as well as increasing the visibility and impact of research[2]. Although open science has not been fully embraced by everyone in the research community, there have been many successful open science projects. By working with researchers and policy makers, as librarians, we can help the open science movement to continue to grow.

...PAST

For much of history, science has been based on observation and description. The development of the scientific method resulted in a shift from description to experiment and measurement. Below is a summary[3] of the evolution of the scientific process:

Empirical science
- Thousands of years ago - describing natural phenomena

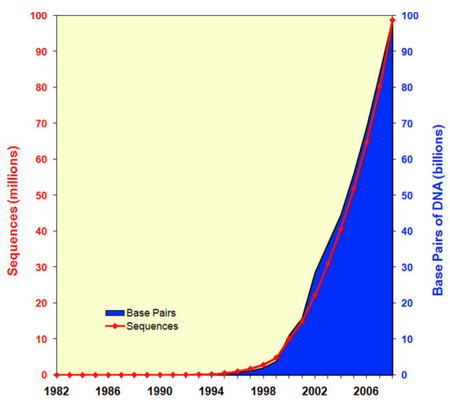
Theoretical science
- Past few hundred years - using models and generalizations to explain phenomena

Computational science
- Past few decades - using digital technology to simulate phenomena

THE FOURTH PARADIGM

The past 10 years or so have seen the most recent evolution in the scientific process, which was coined 'the fourth paradigm' by computer scientist Jim Gray[3]. The fourth paradigm refers to data-intensive scientific discovery, the focus of which is reusing existing data to find new meaning.

THE HGP AND THE BERMUDA AGREEMENT



The Human Genome Project (HGP) officially began in 1990. Not long afterward, labs all over the world began to churn out large volumes of genomic data. The following graph illustrates the exponential increase in data that was deposited in the data repository GenBank between 1982-2008[4].

Despite the positive implications of the HGP on medical advancements, some scientists were reluctant to deposit sequences openly, so in 1996 a number of researchers from various sequencing centers around the world met in Bermuda and set principles for the release of data. Now known as the Bermuda Agreement[4], the principles were:

1. **Primary genomic sequences should be in the public domain**
2. **Primary genomic sequences should be rapidly released**

The HGP and the Bermuda Agreement set a precedent for openness in the field of human genetics and successfully changed scientific culture.

...PRESENT...

Thousands of open science tools are used by researchers every day. Different tools support different functions at various stages of the research process. The following are some examples of the different types of open science tools currently used:

- **Project planning**
  - Tools that track the development of ideas and experimental methods
- **Data collection and analysis**
  - Data repositories and tools that make use of collective intelligence and citizen science
- **Information retrieval**
  - Integrated search tools that link data with literature
- **Demonstrating impact**
  - Networking tools and tools that illustrate impact of all forms of contributions

VARIOUS TOOLS

CURRENT CHALLENGES

Despite the ever-increasing number of tools being developed, a number of barriers continue to prevent open science from being adopted more widely.

Legislation and policy
- International, federal, provincial/state, institutional level
- Intellectual property rights issues

Infrastructure
- Data deluge
- Economic sustainability
- Interoperability

Training and support
- For researchers and information professionals

Scientific culture
- Promotion and tenure is based on the impact of publications
- Fear of misinterpretation

FUTURE...

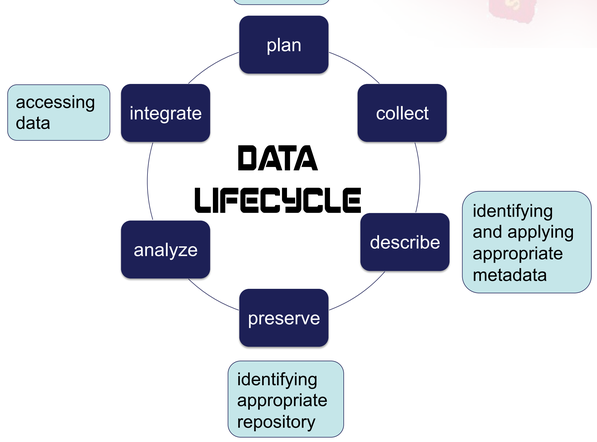
SOLUTIONS

The following solutions proposed by the Royal Society, Science Policy Centre[5] address the challenges mentioned.

Challenge	Solution
Legislation and policy	• standards for sharing information in order to ensure usability and interoperability • data deposit requirements from granting agencies and institutions
Infrastructure	• continual development of new software tools able to hold large data sets • ensuring data are published in a reusable form
Training and support	• involvement of experts (i.e. librarians) able to manage and support digital data
Scientific culture	• increased openness among scientists themselves, as well as with the public and media • consistent recognition for gathering and sharing data

THE ROLE OF THE LIBRARIAN

As librarians, we have an important role in participating in open science. Our various skills can help ensure that data are maintained with the goal of long-term preservation and future use. The figure below illustrates the data lifecycle and the points at which librarians can be involved:



Open science is having a similar impact on scientific progress as the advent of scholarly publication, and for that reason has been called a 'second open science revolution'[5]. As the movement continues to grow, we will be required to deal with larger and larger data sets and be comfortable using dynamic, integrated tools. Promotion and tenure requirements will evolve and consider a wide variety of research contributions. Standards and policies at international, national and institutional levels will ensure interoperability of tools, and granting agencies will increasingly require data to be deposited in open repositories. Until then, as librarians, we can continue to advocate for change and be active participants in open science.

SELECTED SOURCES

1. Gezelter, D. (2011). An informal definition of OpenScience. Retrieved from http://www.openscience.org/blog/?p=454
2. Shearer, K. (2011). *Comprehensive brief on open access to publications and research data for the federal granting agencies*. Retrieved from http://www.science.gc.ca/default.asp?lang=en&n=2360F10C-1
3. Hey, A. J. G., Tansley, S., & Tolle, K. M. (Eds.). (2009). *The fourth paradigm : data-intensive scientific discovery*. Redmond, Wash.: Microsoft Research.
4. U.S. Department of Energy. (2011, Sept. 19, 2011). Human genome project information Retrieved July 8, 2012, from http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml
5. The Royal Society. 2012. Science as an open enterprise. *PDF Rep.*, London, England
6. Nielsen, M. A. (2012). *Reinventing discovery : the new era of networked science*. Princeton, N.J.: Princeton University Press.

Images: The Spanish Cucumber http://royalsociety.org/policy/projects/science-public-enterprise/report/; The Fourth Paradigm http://research.microsoft.com/en-us/collaboration/fourthparadigm/; ResearchGate www.researchgate.net/; Galaxy Zoo www.galaxyzoo.org; myGrid www.mygrid.co.uk; NCBI www.ncbi.nih.gov; Open Notebook Science http://onschallenge.wikispaces.com; FoldIt http://fold.it/portal/; total-impact www.total-impact.org; DataCite www.datacite.org; "gavel" symbol by Connie Shu, "cloud computing" symbol by Andrew Forrester, "lecturer" symbol by Zbigniew Flakus and "nuclear" symbol by the Noun Project al taken from the Noun Project collection (blog.thenounproject.com/).