Initiatives of Research Data Literacy for Data Management in Libraries

P. Ramkumar
Librarian
Srimath Sivagnana Balaya Swamigal Tamil Arts and Science College, Mailam-604304, Tamilnadu
ramkumar_pe@yahoo.co

Kannan
University Assistant Librarian
Thiruvalluvar University, Vellore-632115,
kankishore2012@gmail.com

Abstract

There is a paradigm shift in practicing methods of research which has entered the phase of eScience. The very essence of science is changing, particularly through the deployment of electronic networks and high-speed computers which are the two core components of e-science. This transformation is not limited to the natural sciences, where e-science has become, the modus operandi but also the domains of the social sciences and humanities. Preserving and archiving raw data for reuse has become imminent in the context of time and money. Librarians have the responsibility to initiate research data literacy to facilitate research data management. This article presents an overall picture of research data literacy and role of library professionals. India, in order to promote its scientific productivity needs to join the stream of research data management which needs initiatives from the academic libraries.

Keywords

Research Data; Data Literacy; e-science; Fourth Research Paradigm; Data Management.

Electronic access
The journal is available at www.jalis.in

Journal of Advances in Library and Information Science
ISSN: 2277-2219 Vol. 7. No.1. 2018. pp.128-131

INTRODUCTION

Practicing methods of scientific research have shifted fundamentally during the last several decades. "Enhanced science, e-science, is one of many terms used to describe recent transformations in the scientific enterprise. The term ''e-science'' sometimes refers to ''electronic science,'' but the variant ''enhanced science'' more specifically relates to ''Internet-enhanced'' or ''cyber infrastructure-enhanced'' science[1]. The genealogy of these and other terms was sketched in a keynote address[2].

The Trend today is that many disciplines overlap and use data from other sciences. Internet proves its potential in unifying all literature and data. In a spiraling process, it goes from literature to computation to data back to literature. This has facilitated Information as well as data anywhere, anytime and at any device thus increasing the Scientific Information Velocity. No doubt, there is a huge increase in Science Productivity which is the result of eScience. The term "eScience" was coined in the year 2000 by John Taylor, when he was director general of the UK Research Councils. Taylor recognized the increasingly important role that IT must play in the collaborative, multidisciplinary, and data-intensive Scientific research of the 21st century and "used the term eScience to encompass the collection of tools and technologies needed to support such research"[3].

[4]Federer observed that, "Researchers of the 21st century often rely on large digital datasets, and sometimes they are using data that they themselves did not gather, but that they obtained from public sources for reuse. Researchers in many fields must comply with new policies from funders and journals that require them to share their data or write data management plans, tasks that most researchers have not had to undertake in the past. Whatever a researcher's particular field of study, chances are good that the ways that research is conducted have substantively changed in the last several years. As a result, researchers may find that they need new skills and knowledge to work most effectively and take advantage of the new opportunities that this age of data-driven research presents". The more appropriate skill required to meet the challenges of Big data is Research data literacy.

RESEARCH DATA LITERACY

The term 'Research data literacy' gradually emerged as a mainstream term. It has gained a potential focus

in the progress of the 'Data Revolution', as well as the 'Big Data' discussions. Academics, experts, and policymakers have been considering means and ways of archiving raw data in order to enable the research community make better use of the vast amount of data available to them. A right beginning to achieve this is the concept of Research data literacy. Literacy is a common term used for various purposes. As defined by UNESCO[5] , " Literacy the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society."

DEFINITIONS

Data

Data can be an object, variable, or piece of information that has the perceived capacity to be collected and stored in order to be identified, used and reused. "It comes largely in two forms: structured and unstructured. Structured data are essentially answers to questions asked by the collector of data, are generally easy to organize and identify and have a strict hierarchy that is not easily manipulated (i.e. responses to a survey organized in a table format and information about people's years of education and income in a chart). Unstructured data are not readily amenable to automated analysis and often are used in ways that differ from the intended purpose when collected (such as photos, videos, tweets), and do not need to follow a hierarchical method of identification"[6].

Data Literacy

[7]Deahl defines that, "data literacy is the ability to understand, find, collect, interpret, visualize, and support arguments using quantitative and qualitative data." In the views of Vahey et al.,[8] data literacy includes the ability to formulate and answer questions using data as part of evidence-based thinking; use appropriate data, tools, and representations to support this thinking; interpret information from data; develop and evaluate data-based inferences and explanations; and use data to solve real problems and communicate their solutions.

The Fourth Research Paradigm

The development of science as a method of approach to research problems has passed over three realms to reach the present day fourth research paradigm. Jim Gray[9] described eScience as a "transformed scientific method" or "the fourth paradigm." Research in its life cycle, was originally (i) empirical. During the last few hundred years, (ii) theoretical models emerged. More recently, researchers are able to use computational tools to explore simulations of (iii) complex data environments. Now in the fourth paradigm science, a researcher has access potential to vast quantum of data from experiments and instruments, massive simulations, meta analysis of research results, and much more. Gray argues that this is a new way of doing research and requires a new model for conducting scientific inquiry. Data Science also known as eScience is the set of tools and technologies to support data federation and collaboration for analysis and data mining; for data visualization and exploration and for scholarly communication and dissemination.

Jim Gray, a Technical Fellow and visionary at Microsoft's eScience group originated the phrase 'fourth research paradigm'. There had been precursors in the history of science for the fourth research paradigm. "Johannes Kepler, an assistant to Tycho Brahe took Brahe's catalog of systematic astronomical observations and discovered the laws of planetary motion. This is one of the incidents that marked the distinction between the mining and analysis of experimental data captured and archived and the creation of theories."[10] This is one aspect of the Fourth Research Paradigm – occurred earlier but as a concept identified and named recently.

Till a few decades ago, by convention, the results of research were published as articles or reports while the pertaining data on which scientific theories were based was often buried in individual scientific notebooks or, just stored on magnetic media, never to be accessed and reused. In the views of experts, "The Data Intensive Scientific Discovery or fourth Paradigm Science is more than just using data, analysis, integration and discovery of new principles but ultimately become a "crowd sourced" scientific process". The Fourth Paradigm science or research is an impact of Information and Communications Technology. It is a new method of pushing forward the frontiers of knowledge by means of enabling technologies for gathering, manipulating, analyzing and displaying data.

A current trend recognized within the scientific community is the ongoing deluge of scientific data which is fundamentally transforming almost all areas of academic research. Stake holders of research drawn from scientists and engineers to social scientists and humanities researchers – now require tools, technologies, and platforms that seamlessly integrate into standard scientific methodologies and processes. "The Fourth Paradigm" refers to the data management techniques and the computational systems needed to manipulate, analyze, visualize, and manage large amounts of research data.

Trend setting

According to the National Science Foundation (NSF) Data Sharing Policy 2010, it has become mandatory for the "Investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing". In addition, funding agencies insist on a two paged data management plan along with the research proposal.

Data Scientists

"In 2005, the National Science Board of the National Science Foundation published "Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century," which began a dialogue about the importance of data preservation and introduced the issue of the care and feeding of an emerging group they identified as "data scientists": The interests of data scientists—the information and computer scientists, database and software engineers and programmers, disciplinary experts, curators and expert annotators, librarians, archivists, and others, who are crucial to the successful management of a digital data collection—lie in having their creativity and intellectual contributions fully recognized." [11]

Initiatives for Library Data Services

"Data services for students and faculty have existed in research libraries for decades, but it was the rise of computational research in the sciences and engineering -- and the data deluge that followed -- that led to the development of research data management services, defined as the storage, curation, preservation, and provision for continuing access to digital research data", in the views of Hey and Trefethen[12] ; Lewis[13] . Data intensified Science compels the library community resort to Research data management.

There are a few steps that the library professionals have to initiate towards Research data management. The situation requires the library community foster the development of generic Laboratory Information Management Systems (LIMS); Foster research into scientific data management, data analysis, data visualization, and new algorithms and tools; Establish and extend digital libraries that can support other sciences in the same way the National Library of Medicine supports the bio-sciences. The most significant of all is that they have to foster the development of new document authoring tools and publication models as well as foster the development of digital data libraries that contain scientific data (not just the metadata) and support integration with published literature.

In the context of the emerging rapid growth of eScience and the use of large datasets in research, the U.S. Office of Science and Technology Policy issued a mandate in 2013 designed to expand public access to the results of federally funded research. It specified explicitly that in practical terms, "results" refers to both published articles and their associated research data. Consequently, research data management has steadily developed into an area of new opportunities and roles for libraries. Many libraries that serve research communities have begun to develop services to support researchers' data-related needs.

CONCLUSION

Science was very much discipline-centric decades ago. Today, Science has entered the fourth paradigm due to significant advances that are being made as a result of multidisciplinary collaboration and the advent of Big data and thus data intensified science. Braxton et al.[14], tipped on the pre-requisites for data archiving libraries stating that,"The Library's public-facing eResearch support services cannot advance substantively unless there is both campus-level technical and personnel infrastructure to support data curation and preservation and domain-specific expertise to support research data management, preservation, and sustained access". All research oriented institutions in general, academic institutions in particular that are in receipt of their funding should develop a clear roadmap for research data management which is fundamental to research data use and reuse. The research data landscape is complex as it is constantly evolving, making it a challenge for libraries and librarians to stake out their territory. Experts believe that, "examining the details of the landscape with a finer eye, particularly within

individual disciplines such as veterinary medicine, can provide valuable information for institutions developing a research data service". Library professionals should equip their knowledge both in research data literacy as well as research data archiving in order to implement research data effectively for an efficient and effective recall every unit of data with precision. In the opinion of experts, library professionals only on their understanding "how researchers interact with their data at various points in the life cycle, as well as recognizing pain points for researchers, can help librarians develop targeted services to meet the particular needs of researchers at their institutions". India, in order to promote its scientific productivity, in a highly competitive world, needs to join the stream of research data management which needs initiatives as well as support from the academic libraries.

REFERENCES

[1]. Jankowski, Nicholas W., Exploring e-Science: An Introduction Journal of Computer-Mediated Communication, 12 (2007) pp. 549–562.
[2]. aniel Atkins, held at the Network for Earthquake Simulation (NEES) World Forum in March 2006.http://www.nees.org/About_NEES/Announcements/presentations/NEESWorldForum/Keynote-Intro/Atkins/NEES3-17-06.pdf
[3]. John Taylor e-Science definition: http://www.e-science.clrc. ac.uk/.
[4]. L. Federer / Research data management in the age of big data: Roles and opportunities for librarians. Information Services & Use 36 (2016) 35–43 35 DOI 10.3233/ISU-160797.
[5]. [5]Bhargava, Rahul., et al., Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data, September 2015, Working Paper for Discussion, MIT center for Civic media.
[6]. "Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data." Data-Pop Alliance White Paper Series. Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute) and Internews. September 2015.
[7]. Deahl, E. (2014). Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments. M.S. Thesis, Massachusetts Institute of Technology.
[8]. Vahey, P., Yarnall, L., Patton, C., Zalles, D., and Swan, K. (2006). Mathematizing middle school: Results from a cross-disciplinary study of data literacy. In Annual Meeting of the American Educational Research Association, San Francisco, CA.
[9]. Gray, Jim. "Jim Gray on eScience: A transformed scientific method." In The fourth paradigm:Data-intensive scientific discovery, edited by Tony Hey, Stewart Tansley and Kristin Tolle, xix Microsoft, 2009.
[10]. Hey, Tony., Tansley, Stewart., and Tolle, Kristin., Ed., The Fourth Paradigm, : Data-Intensive Scientific Discovery, Washington: Microsoft Research at http://research.microsoft.com
[11]. Hey, Tony, and Anne Trefethen. "The Data Deluge: An e-Science Perspective." In Grid Computing: Making the Global Infrastructure a Reality, edited by F. Berman, G. Fox and T. Hey, 809-824. Chichester: Wiley, 2003.
[12]. Lewis, Martin. "Libraries and the Management of Research Data." In Envisioning Future Academic Library Services Initiatives, Ideas and Challenges, edited by Sue McKnight, 145-168. London: Facet Publishing, 2010.