International Journal of Advanced Smart Convergence Vol.12 No.4 98-103 (2023)
http://dx.doi.org/10.7236/IJASC.2023.12.4.98

IJASC 23-4-11

Changes in Research Paradigms in Data Intensive Environments

Minsoo Park

Department of Data Science, School of Applied Artificial Intelligence Engineering,
Kangnam University
mspark7@gmail.com

Abstract

As technology advanced dramatically in the late 20th century, a new era of science arrived. The emerging era of scientific discovery, variously described as e-Science, cyberscience, and the fourth paradigm, uses technologies required for computation, data curation, analysis, and visualization. The emergence of the fourth research paradigm will have such a huge impact that it will shake the foundations of science, and will also have a huge impact on the role of data-information infrastructure. In the digital age, the roles of data-information professionals are becoming more diverse. As eScience emerges as a sustainable and growing part of research, data-information professionals and centeres are exploring new roles to address the issues that arise from new forms of research. The functions that data-information professionals and centeres can fundamentally provide in the e-Science area are data curation, preservation, access, and metadata. Basically, it involves discovering and using available technical infrastructure and tools, finding relevant data, establishing a data management plan, and developing tools to support research. A further advanced service is archiving and curating relevant data for long-term preservation and integration of datasets and providing curating and data management services as part of a data management plan. Adaptation and change to the new information environment of the 21st century require strong and future-responsive leadership. There is a strong need to effectively respond to future challenges by exploring the role and function of data-information professionals in the future environment. Understanding what types of data-information professionals and skills will be needed in the future is essential to developing the talent that will lead the transformation. The new values and roles of data-information professionals and centers for 21st century researchers in STEAM are discussed.

Keywords: *Big Data, Scholarly Communication, Open Access, eScience, Nanotechnology, STEAM*

1. Introduction

In September 2005, the U.S. National Science Board publised Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century [1]. This report addresses the new roles and responsibilities of individuals and institutions in data intensive environments. The individuals or institutions involved are data authors, data managers, and data scientists. These roles are considered as follows. “Authors of data” are scientists, educators, and students in the domain who are very interested in research that arises from data. “Data managers” are information engineers, computer scientists, and information scientists who are responsible for calculating, accumulating, and accessing data for analysis. “Data scientists” are responsible for

Manuscript Received: october. 10, 2023 / Revised: october. 16, 2023 / Accepted:october. 21, 2023
Corresponding Author: mspark7@gmail.com
Associate Professor, Department of Data Science, Kangnam University

creating/collecting, processing, analyzing, and producing solutions/insights to better understand the rapidly changing phenomena around us in the big data environment.

As technology advanced dramatically in the late 20th century, a new era of science arrived. The emerging era of scientific discovery, variously described as e-Science, cyberscience, and the fourth paradigm, uses technologies required for computation, data curation, analysis, and visualization. The emergence of the fourth research paradigm will have such a huge impact that it will shake the foundations of science, and will also have a huge impact on the role of data-information infrastructure. In the digital age, the roles of data-information professionals are becoming more diverse. In this paper, we aim to understand changes in research paradigms in data-intensive enrionments and to review and redefine the value and role of data-information professionals and centers for 21st century researchers in the STEAM field.

2. Changes in Research Paradigm: Data-Centered Research Paradigm

First used by John Taylor, the UK's Director of Science and Technology in 1999, eScience refers to large-scale science conducted through global collaboration on the Internet [2]. According to the UK's National e-Science Centre, e-Science, a collaborative science enterprise, requires individual scientists to have access to large data, large-scale computing resources, and high-performance visualization [2]. All fields of natural science (biology, chemistry, physics, etc.) as well as biomedical and social sciences are included in eScience. The UK's eScience program consists of a variety of resources, centers and personnel. There is a National e-Science Center run by the University of Glasgow and the University of Edinburgh. There are eScience Centers in each region that support universities and projects.

The eScience initiative has been actively promoted in the UK since 2000. In the United States, various eScience activities are being carried out at the national level under the term cyberinfrastructure. The term cyberinfrastructure was first used in 2003 by the U.S. National Science Foundation and Blue Ribbon Committee [3]. This is a term that emerged as NSF, a major government agency that supports basic research, sought ways to benefit all scientists, engineers, researchers, and citizens by removing existing barriers to rapidly evolving high-performance computing [3].

Cyberinfrastructure refers to a research environment that supports data collection, data storage, data management, data integration, data mining, data visualization, and other computing and information processing services on the Internet without time and space constraints. From a scientific perspective, cyberinfrastructure can be viewed as a technological solution to efficiently support laboratory space, data, computers, and researchers' research activities.

Cyberinfrastructure projects in the United States are primarily supported by the National Science Foundation and the Department of Energy's Office of Science [3]. NSF's Office of Cyberinfrastructure supports the TeraGrid Project, which provides integrated resources and services operated by supercomputing centers in the United States [3].

As eScience emerges as a sustainable and growing part of research, data-information professionals and centeres are exploring new roles to address the issues that arise from new forms of research. The functions that data-information professionals and centeres can fundamentally provide in the e-Science area are data curation, preservation, access, and metadata. Basically, it involves discovering and using available technical infrastructure and tools, finding relevant data, establishing a data management plan, and developing tools to support research. A further advanced service is archiving and curating relevant data for long-term preservation and integration of datasets and providing curating and data management services as part of a data management plan.

The National Center for Biotechnology (NCBI) was established in 1988 [4]. NCBI exists to plan, develop, implement, and manage automated systems for collecting, storing, retrieving, analyzing, and distributing information related to molecular biology, biochemistry, and genetics. In terms of international cooperation, a representative example is the International Nucleotide Sequence Database Collaboration. This joint collaboration involves GenBank in the United States, EMBL in Europe, and DNA Databank in Japan, and is a huge model in that it provides data backup and accessibility. And this is international data sharing at its best. The more data and connections there are, the more they will be used, and the more data will flow in.

Although the field of eScience (sometimes called cyberinfrastructure) is recent, it has brought significant changes in the way researchers in science and technology conduct research. Instead of conducting experiments or tests involving vast numbers of variables, scientists today can test principles of biology, chemistry, and physics, through computer simulations. eScienc is implemented through distributed, global-scale collaboration, and increasingly through the Internet. The characteristics of such joint (collaborative) research are, first, that it requires access to very large data sets and requires very large-scale computing resources. And for the individual user, the scientist, the results produced by the computer need to be displayed to a significant extent visually.

From an international perspective, many countries and international organizations have already recognized the social, economic, and scientific benefits that can be gained from open access to scientific and technological digital data. In 2004, more than 30 countries, including the United States, issued a joint statement aimed at establishing common access to digital research data produced with public funds. Because the international exchange of scientific data, information, and knowledge greatly increases the scope and scale of research, and the resulting impact, these countries are jointly defining the implementation steps necessary to operate the global science and technology system. At the institutional level, universities are developing approaches for digital data archiving, curation and analysis. Each institution shares best practices to develop a digital library that collects, preserves, indexes, and shares research materials produced by professors and researchers. The technical implementations of these systems are mostly open source and support interoperability with each other.

3. Scholarly Communication in Data-Intensive Environments

Research institutions have made a significant contribution to the academic communication cycle by collecting, organizing, and managing works published by researchers and ensuring their long-term use. Recently, research institutions are developing publishing services as well as converting existing analog content into digital form. Research institutions expand from existing areas and move into other areas in the academic communication cycle. Rather than holding on to the end of publication, they are getting closer to the research process and the academic communication that occurs throughout the research process.

Open access has had a profound impact on the role of academic libraries. Libraries are changing from the role of gatekeepers guiding information produced elsewhere to actively supporting the process of producing and distributing scientific and technological information. Libraries must build new capabilities and develop technology platforms for publishing and archiving. KPubS [5] is an international academic journal open access publishing platform that aims to build an infrastructure that can manage and process domestic English-language academic journals in a full cycle, from paper submission to review, editing, publication, data storage, web service, and international dissemination. Figure 1 shows the overview of KPubS model. The purpose of each stage is as follows: the creation stage is to secure excellent papers at an international level, the archiving stage is to maintain standard management of data, the web service stage is to build and operate online service pages, and the circulation stage is to increase international exposure and support international indexing registration.

Lor explains why it is important to choose open access [6-7]. Lor points out that the open access movement emerged as a result of a convergence of economic crisis, moral crisis, and enabling technology. An economic problem is the price of academic journals, which rises every year. It is becoming a norm to cancel subscriptions to academic journals every year, and book purchase fees must be drastically reduced in order to subscribe to what is considered a core academic journal. For developing countries, this situation is even more serious. The second element, moral crisis, has two issues. The first issue is that researchers in developing countries cannot obtain the necessary access to scientific and technological information.

The digital divide is not limited to the technology divide, but also applies to the content gap and knowledge gap. The second issue is that there is a loss of balance and unfairness in the relationship between authors, journal publishers, and users. Authors wish to submit their research results free of charge, and editors and referees wish to peer-review the research results without receiving financial compensation. Scholars publish their research in peer-reviewed journals to gain professional recognition. Costs paid by academic journal subscribers arise during the publication and distribution process. The third factor is the advent of the Internet. Digital technologies are used to control or improve accessibility. It appears that digital technologies can provide significant cost savings by reducing not only set-up costs but also marginal costs. Old traditions, new technologies, and new roles merged to create unprecedented public properties. The long-standing tradition is that scientists and scholars seek to publish their research results in academic journals for free. The new technology refers to the Internet, and the new role refers to the library's rebirth as a publisher of academic information, which is a public property.

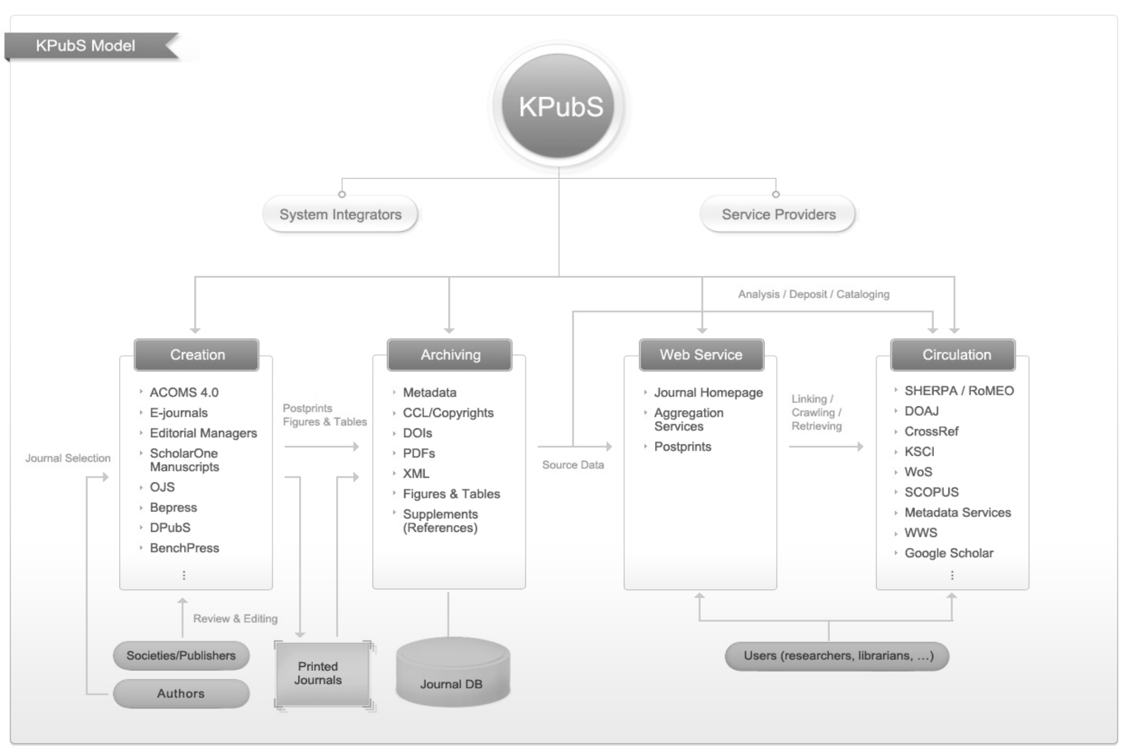

Figure 1. KPubS Model

Nanotechnology is a field of research that continues to develop at the latest and requires considerable interdisciplinary cooperation. For learning and discovery in the field of nanotechnology, information exchange, data sharing and accumulation, and archiving are important. NanoHUB [8] is a major project supported by the Network for Computational Nanotechnology (NCN). Financial support was provided by the National Science Foundation (NSF). Although this case does not involve the accumulation and management of large data sets, it is notable as an information exchange case.

This is no different from what has now been thought of as the role of data-information centeres. Interesting insights can be gained from the role that NanoHUBs play when collecting and sharing nanotechnology information. What is the current and future role of centers such as NanoHUBs? Who is responsible for preserving records of scholarly communication resulting from information and research exchanges such as NanoHUBs? First of all, to understand the driving force of nanotechnology, it is helpful to reexamine its definition. On NanoHUB's home page[8], nanotechnology is defined as follows: a nanometer is one billionth of a meter. In other words, it is 25,000 times thinner than the width of a human head, 200 times smaller than a typical virus, and is the size of 3 to 4 atoms from corner to corner. Therefore, it is important for researchers around the world to collaborate to study the properties of nanometers and their applications in science. Nano Hub's mission statement is written on the website as follows: NanoHUB, a web-based resource for research, education, and collaboration in nanotechnology, is an initiative of the NSF-funded Network for Computational Nanotechnology (NCN).

NCN is a network of universities with a pioneering view in developing nanotechnology from science to manufacturing through innovative theory, experimental simulations, and new cyberinfrastructure. NCN's students, staff, and university faculties are developing the Nanohub Science Gateway while also using the Nanohub for their own research and teaching. The Nano Hub, a resource for the entire Nano community, focuses on NCN's three themes: nanoelectronics, NEMS/nanofluidics, and nanomedicine/biology. Scientists, computer scientists, and applied mathematicians work together to propose solutions to various problems. NonoHUB is a major place for information exchange.

NanoHUB allows simulations using nanoelectronics, NEMS/nanofluidics, Nanobio Plus, and other development tools. NanoHUB supports research through links to online seminars, provides links for collaboration or online meetings, and facilitates user groups [8]. Through its teaching and learning section in the field, NanoHUB also provides links to online courses to help university students (and researchers) study nanotechnology. It also provides textbooks on nanotechnology for children.

The scholarly communication facilitated by NanoHUB is important for this newly defined field of research. Without the exchange of research findings, simulation modeling and integration into the curriculum, and student learning from elementary school to college, there would not have been the progress currently being demonstrated in the field of nanotechnology. Traditional forms of scholarly communication, such as research publications, would not be able to meet the real-time needs of nanotechnology researchers. Furthermore, NanoHUB can deliver information through new forms of media, such as interactive oral presentations or podcasts that can be accessed via iTunes.

4. Discussion and Conclusion

eScience aims to increase research productivity by connecting computing systems, data storage systems, data repositories, visualization environments, and researchers using software and high-performance networks. eScience conducts research in terms of researchers, data, information, and tools. The vision is to build an interactive, functionally complete, and more ubiquitous digital research environment for the community.

Data-information professionals must move from reading rooms that simply provide information to research centers. In other words, there is a need to move from traditional methods to a more open environment where we can experiment and collaborate with experts. To achieve this, researhers must be able to access data continuously without interruption and continue to use and utilize the data during the whole proess of resaerch. This requires collaboration, inter-agency approaches, infrastructure and integration of collections.

Following changes in research paradigms in these academic fields, there are new roles for data-information professionals in data-intensive research environments: (1) efficiently storing, preserving, creating metadata, and providing access to data; (2) providing linkage functions between publications and data, and between data and scientific research activities; (3) providing related policy and content management consulting services; (4) expert knowledge on open access and open data issues, copyright, and data policy management is required; (5) more interactive cooperation with researchers throughout the entire research cycle (from data generation to provision); (6) performing the role of a data scientist through local and global cooperation and communication; (7) rebalancing (a rebalancing act): local and global cyberinfrastructure; (8) Cultivating new skills and abilities that the world of data requires from information professionals

Adaptation and change to the new information environment of the 21st century require strong and future-responsive leadership. There is a strong need to effectively respond to future challenges by exploring the role and function of data-information professionals in the future environment. Understanding what types of data-information professionals and skills will be needed in the future is essential to developing the talent that will lead the transformation.

References

[1] US National Science Board (NSB). Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century, NSB-05-40

[2] National e-Science Centre, eScience, http://www.nesc.ac.uk

[3] National Science Foundation, Cyberinfrastructure. http://nsf.gov/pubs/2007/nsf07601/nsf07601.htm

[4] Purdue Ionomics Information Management System (PiiMS). Purdue University. http://center.e-enterprise.purdue.edu

[5] KPubS, www.kpubs.org

[6] P. J. Lor, “Libraries in times of open access,” *Information Services and Use*, Vol. 27, No. 4, pp. 193-203, 2007.

[7] P. J. Lor, “Is a knowledge society possible without freedom of access to information?,” *Journal of Information Science*, Vol. 33, No. 4, pp. 387-397, 2007.

[8] NanoHUB, Network for Computational Nanotechnology, http://www.nanohub.org