Chapter 11
Research Data Management:
Models, Challenges, and Actions

Yuhong Cui
Beijing Institute of Technology, China

ABSTRACT

Research Data is acknowledged as constituting elements of the big data landscape and currency of science. With the rise of the open science movement and data sharing policies, research data management has been an emerging practice among domain-specific science and library and information science, while academic libraries have become deeply involved in the development of principles and best practices for managing data for long-term use. Research data management practices in China have a much shorter history and face more challenges. This paper is to explore the development of research data management in China across different disciplines. A conceptual framework of Research Data Management for Development (RDM4D) is proposed to demonstrate the status of data sharing policies, data repositories and libraries actions within three dimensions of societies, technology and humanity in China. Our goal is to bridge the gap between China and other countries and explore which are the proper actions for Chinese academic libraries.

INTRODUCTION

'Big Data,' popular words on the covers of *Science*, *Nature*, *The Economist* and *Wired* magazines, increasingly becomes an emerging topic through the global scholarly and entrepreneurial communities. With the e-Science development and data-intensive research paradigms emerging, research data, linked data, the web of data and open data are referred to as constituting elements of the big data landscape (Shiri, 2014; Borgman, 2014).

As the currency of science and essential to scientific productivity, collaboration, and to discovery itself, the sharing of research data has long been a commonly acknowledged concept among many research communities (Gold, 2007). Borgman (2007) provides four rationales for sharing research data,

DOI: 10.4018/978-1-5225-0550-1.ch011

namely "to (a) reproduce or verify research, (b) make results of publicly funded research available to the public, (c) enable others to ask new questions of extant data, and (d) advance the state of research and innovation."

The rise of the open data and open science data movements, in conjunction with the increasing implementation of data management and sharing policies by funding bodies, governments, and journals, has led to an explosion in the number of research articles and proceeding papers associated with research data practices and services created to serve institutions, association members and research communities in the United Kingdom, the United States, Australia, Germany and other countries. Meanwhile, data curation and research data management has been the key topics in the ACM/IEEE Joint Committee digital libraries and e-Science international meeting. The Research Data Alliance created in 2013 has grown into a community with more than 2,800 members from over 90 countries who are dedicated to building the social, organizational and technical infrastructures needed to reduce barriers in data sharing and exchange and are also dedicated to accelerating data-driven innovation worldwide.

Academic libraries, with their long experience with information organization and documentation, represent a natural organizational home for data management services and become deeply involved in the development of principles and best practices for managing data for long-term use. Case studies began to emerge in 2008 when Library Trends published a special issue on institutional repositories that described library efforts to include research datasets in their archiving at Johns Hopkins University, Purdue University, and the University of Minnesota (Choudhury, 2008; Delserone, 2008; Witt, 2008). Further evidence of the mainstreaming of research data management in the library field is shown by the publication of books on the subject aimed specifically at library and information professionals by UK and US researchers (Pryor, 2014; Ray, 2014).

Research data management practices in China have a much shorter history, beginning when some senior researchers called for the enhancement of research data sharing in 1994. One of the most important initiatives is the National Scientific Data Sharing Program (SDSP) launched by the Ministry of Science and Technology (MOST) in 2002 (Xu, 2003), later called the National Scientific Data Sharing Infrastructure of China (SDSIC). Unfortunately, most of the datasets conducted from the public funded research programs by universities are still not available to be shared. China ranks second in the production of scientific articles according to SCI database and is becoming a leading scientific data producer. Research data management faces more challenges and is still very new to both the Chinese data scientific community and library and information science.

This chapter aims to examine the development of research data management in China across the disciplines in order to bridge the gap between China and other countries and explore which are the proper actions for Chinese academic libraries.

In the following sections, firstly a model framework of Research Data Management for Development (RDM4D) is demonstrated after the systematic literature review by bibliometric analysis across the domain-specific science disciplines. Then, some examples of data centers and repositories are detailed for infrastructure of research data management. We next look at the national and organizational policies in the subsequent section, specifically compared to the funder policy expectation and regulation between NSFC in China, NSF in the United States and ESPRC in the UK. Last but not least, we review the actions taken by Chinese academic libraries and discuss how to promote the best practice on an institutional level under the Chinese research culture.