The Journal of Academic Librarianship xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

The Journal of Academic Librarianship

journal homepage: www.elsevier.com/locate/jacalib

Survey on the Needs for Chemistry Research Data Management and Sharing

Xiujuan Chen[a,c], Ming Wu[b,*]

[a] Chengdu Library and Information Center, Chinese Academy of Sciences, No. 16, South Section 2, Yihuan Road, Chengdu, Sichuan 610041, PR China
[b] National Science Library, Chinese Academy of Sciences, No. 33 Beisihuan Xilu, Haidian District, Beijing 100190, PR China
[c] University of Chinese Academy of Science, No. 19A, Yuquan Road, Shijingshan District, Beijing 100049, PR China

ARTICLE INFO

Keywords:
Chemistry
Data management and sharing
Data management lifecycle
Needs survey
Library data service

ABSTRACT

This paper aims to reveal the situation of research data in chemistry research process and chemistry researchers' need for data management support from five perspectives, i.e., data generation and collection, data recording and processing, data preservation and backup, data publication and sharing, needs for data management and sharing services. Our survey is based on a questionnaire carried out among 119 subjects, i.e., researchers and graduate students in chemistry of Chinese Academy of Science. The analysis results provide us with a better understanding on the current attitudes and needs of researchers and graduate students about data management and sharing in chemistry. Although this survey was implemented in chemistry, it could provide us with some inspirations for designing a range of library services for other disciplines, particularly in promotion, consulting and training of research data management and sharing, and research data storage.

Introduction

In the context of e-science and open data, some governments, international organizations, funding agencies and periodical publishers have issued a series of research data management and sharing (RDMS) policies in recent years. Researchers are encouraged or required to share research data along with their research outputs. These open data policies have actively promoted research data sharing and reusing during research data lifecycle.

In reality, researchers not only have different attitudes and expectancies for managing and sharing research data (Carlson & Stowell-Bracke, 2013; Hall, 2013; Huang et al., 2012; Tenopir et al., 2011), but also meet many challenges in management data especially for different disciplines of small science (Borgman, 2012; Borgman, Wallis, & Enyedy, 2007; Gray et al., 2005). During the past few years, issues relating to research data management in general and libraries' data management services for researchers in particular, have attracted a great deal of attention in the library community, so that a number of guides or handbooks (e.g., The Medical Library Association Guide to Data Management for Librarians, Data Management: A Practical Guide for Librarians, Research Data Management: Practical Strategies for Information Professionals, etc.) have been published. However, these practices are not adequate to inform libraries in their design of data management services for researchers. To make sure that libraries can provide personalized, specific and effective services for researchers, it is necessary to understand the current situation of research data management and needs for RDMS services in the research process. Therefore some studies were conducted to survey on the current states and needs of researchers and graduate students for RDMS(Anderson et al., 2007; Carlson, 2012; Peters & Dryden, 2011), and others are increasingly engaged in understanding practices and developing new services to satisfy research data management requirements, particularly within academic library(Bardyn, Resnick, & Camina, 2012; Kim, 2015; Williams, 2013).

National Science Library (NSL) is an academic library attached to Chinese Academy of Sciences (CAS). In recent years, NSL has initiated subject librarian system and set up full time subject librarian's team with diverse subject backgrounds. Providing subject information services for the researchers and graduates of the institutes of CAS in different research fields are one important aspect of subject librarians' job. With RDMS becoming more and more important, many researchers also have shown much concern about the issue. So in this study, we conducted the survey for chemistry researchers and graduates in Chemistry Institute of CAS from five perspectives of research data in chemistry research process, i.e., data generation and collection, data recording and processing, data preservation and backup, data publication and sharing, needs for data management and sharing services. By means of questionnaires survey, we collected solid evidence and raw data on RDMS in chemistry. Based on the statistical analytical results, we have a better understanding on their current attitudes and needs of researchers about their data management and sharing. Although the survey was implemented in chemistry, it could be a reference for us to design a

* Corresponding author.
E-mail address: wum@mail.las.ac.cn (M. Wu).

http://dx.doi.org/10.1016/j.acalib.2017.06.006
Received 21 March 2017; Received in revised form 14 June 2017; Accepted 14 June 2017

range of library services for other disciplines to satisfy researchers' needs for data management and sharing around data lifecycle, particularly in promotion, consulting, training and storage of chemistry research data.

Methodology

Questionnaire design

Preliminary preparation of the questionnaire began in July 2015. After some comprehensive revisions, the second draft of the questionnaire was formed in early September. Then the pre-test of questionnaire was carried out with ten random graduates of Shanghai Institute of Organic Chemistry (SIOC), CAS, to test whether the questionnaires were easy to understand or not. Based on the feedbacks of the graduates, we adjusted the questionnaire precisely and formed the final questionnaire draft of "Survey on the needs for chemistry research data management and sharing".

The questionnaire includes a prologue and a survey subject:

(1) Questionnaire prologue: this part describes the background and purpose of the survey, which is to gather information on the status of research data management in the research process and the needs for RDMS services, aiming to carry out the library services of RDMS and support the scientific research.
(2) The survey subject is divided into two parts: personal information of respondents and the main contents of the questionnaire.

① Personal information of respondents includes institute, research field, status, and type of research field. Respondents need to fill in the blank with institute and research field. The respondents are divided into three statuses of the "researcher", "manager", and "graduate students". Research fields are divided into two kinds, that is, the theoretical research and the empirical research.

②The main contents of the questionnaire: Based on research data lifecycle, the main contents of the questionnaire we designed include four parts, "data generation and collection", "data recording and processing", "data preservation and backup", "data publication and sharing" and "needs for data management and sharing services" (shown in Table 1).

Respondents

The survey was conducted from September 2015 to January 2016. Taking into account of the convenience of questionnaire distribution and recycling, as well as representative of the respondents, we surveyed 129 respondents randomly from Changchun Institute of Applied Chemistry (CIAC), Chinese Academy of Sciences (CAS) and Institute of Chemistry Chinese Academy of Sciences (ICCAS). 129 questionnaires were distributed face to face and through a network platform (i.e., the questionnaire star) and the return ratio was 100%. After eliminating the invalid and blank questionnaires, we acquired 119 effective questionnaires (as shown in Table 2).

Table 2
Information of respondents.

	No. of respondents
Institution	
CIAC, CAS	56
ICCAS	63
Status	
Researchers	37
Managers	5
Graduate students	77
Research type	
Theoretical research	20
Experimental research	99
Research field	
Polymer chemistry	37
Physical chemistry	26
Analytical chemistry	19
Material chemistry	13
Chemistry	9
Organic chemistry	8
Inorganic chemistry	5
Applied chemistry	1
Chemical biology	1

Statistical method

We use excel for statistical analysis of each questions and find concentration and dispersion trends of data by frequency and percentage descriptive statistics. The percentage is the number proportion of total respondents.

Results

Data generation and collection

This part intends to get an overview of data generated by respondents, including data types, formats, frequency and sizes. Both journals and data repositories make certain claims on data types, formats and scale, and understanding these basic characteristics is helpful for researchers to publish data smoothly in the late stage. Understanding the updates of research data can facilitate tracking the changes of research data constantly on later.

Types of data generated during research

Data types that respondents' research projects often generate are various, as shown in Fig. 1. The most common type of data generated by respondents is experimental data (79.83%), followed by observation data (36.13%); the less common data types, by contrast, are reference data and simulation data. Experimental data is reproducible, however, the cost tends to be higher; observation data can be generated in real time and it is unrepeatable and irreplaceable. What's more important, these two types of data are worth preserving as the first-hand data of research projects. If these data cannot be preserved properly, it will impact on the research process and data utilization in terms of finance cost or time cost.

Table 1
Questionnaire design framework.

No.	Questionnaire design	Content	No. of questions
1	Data generation and collection	Investigating chemistry research data types, formats, frequency and sizes	Q1—Q4
2	Data recording and processing	Investigating the ways of recording research data, tools and software to process research data	Q5—Q6
3	Data preservation and backup	Investigating data preservation, data lost, data management requirements, ways of managing research data	Q7—Q10
4	Data publication and sharing	Investigating the situation of respondents publishing data and recognition to journal data requirements and data repository	Q11—Q15
5	Need for data management and sharing services	Investigating the demand degrees, the services needed most, the best way for providing services, their suggestions to RDMS services	Q16—Q20