Future Directions of Academic Libraries in Data Curation

Hui Lyu
Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign

GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE
The iSchool at Illinois

Introduction

Many scientists and students are currently working on digital materials and recording their thoughts in digital documents. Therefore, the role of collections of digital resources are enhanced gradually. However, the large amount of intellectual output is usually unorganized and cannot be collected and preserved for a long time. To solve this problem, libraries are the institutions that could best help researchers to disseminate, preserve and manage those data systematically, and to offer data curation service (Heidorn, 2011).

Digital preservation and data curation

Stewardship of digital resources involves both preservation and curation. Digital preservation is mainly about the storage of data based on classification and collection. While data curation aims to make selected data accessible, usable, and useful throughout its lifecycle. Data curation subsumes digital preservation; and digital preservation is the precondition of data curation (Giarlo, 2013).

Challenge of academic libraries

Academic libraries have an opportunity to extend their established digital curation services to the research enterprise (Giarlo, 2013). However, librarians hardly know current data management practices of scholars. Though university libraries are setting about working on campus scholarly data services and research projects, there still exists large space of technical and administrative improvement for academic libraries (Choudhury, 2008). Therefore, the study on future directions of academic libraries in data curation is necessary.

Existing Problems

The emerging role of academic libraries in digital preservation and data curation begins to get more attention in the 21st century. In its short-term development, several issues are coming out correspondingly. There are some existing problems:

- Data management system is not well-designed to preserve intellectual outputs with vulnerable copies.
- Research data can not be permanently preserved under unstable backup programs.
- Metadata and embargoed files are not intelligently connected to the PDF files with immature policies.
- The relationship between researchers and librarians are not established to implement data curation.

Current Measures

Nowadays, academic libraries have adopted several measures to facing those problems. Following are some representative actions and organizations.

Digital Curation Centre Lifecycle Model

A number of projects have developed curation lifecycle models, including the popular Digital Curation Centre Lifecycle Model proposed by Higgins in 2008. The model is shown in Figure 1.

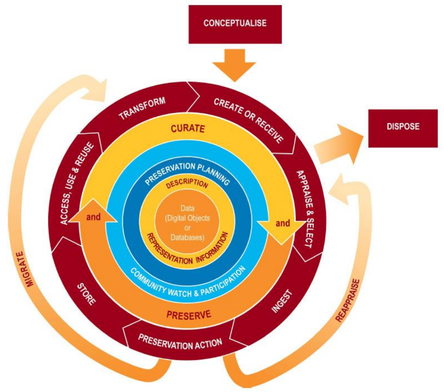

Fig1: DCC Curation Lifecycle Model (Giarlo, 2013).

The DCC Model consists of two parts. The inner rings are full-cycle activities made by data curators in the libraries. The outer rings are sequential actions which need the collaboration between both librarians and researchers.

Institutional Repositories (IR)

A university-based institutional repository offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members (Choudhury, 2008). For instance, UC San Diego has made a relevant program whose logo is shown in Figure 2 (UC San Diego Library).

Research Data Curation Program

Fig2: UC San Diego Library's Research Data Curation Program.

Current research of IR focuses on:

- Providing long-term digital preservation.
- Developing new roles and relationships between libraries and academic communities through the training of "data scientists" or "data humanists".
- Engaging faculty in discussions about the value of IR services to the current scholarly publishing system.
- Making IR an important component in data curation.

Future Directions

To better achieve data curation for academic libraries, future directions of development are necessary to be listed. Based on previous study in this field, several prospects in data curation are demonstrated in Figure 3.

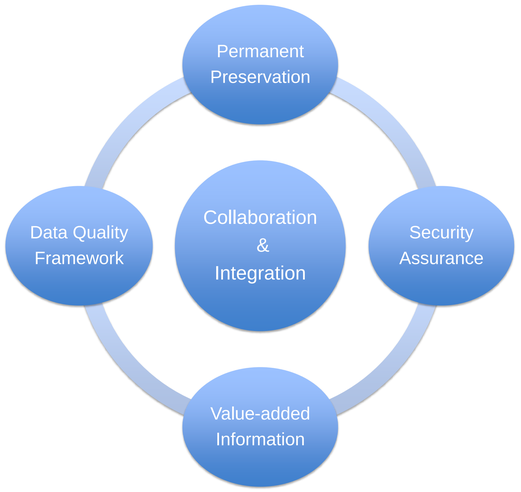

Fig3: Aspects of future directions.

Five aspects together constitute the future directions of academic libraries in data curation.

Collaboration & Integration

Collaboration and Integration is the central aspect and the essential action to be adopted. It includes three parts:

- Collaboration between diverse repositories and organizations of academic community.
- Collaboration between data curators and scientists or researchers.
- Integration of documentation, metadata and valuable context in the data management system.

Data curation is not only a work belonging to librarians, but also should be an after-thought action as one step in the workflows of experts or researchers. Communications of interdisciplinary research domains are needed to truly understand the data and make data curation effectively.

Permanent Preservation

Given the large amount of data, permanent preservation is the foundation and should be noteworthy. It includes:

- Adopting precautionary measures.
- Making multiple backups in different formats and locations to achieve reliable recovery.

Data Quality Framework

Data quality is a concept with multiple dimensions which primarily represent a series of desirable characteristics for an information resource (Curry et al., 2010).

To be trust is the final objective, including authenticity, understandability, usability and integrity (Giarlo, 2013), as is shown in Figure 4.

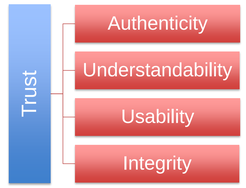

Fig4: Data quality framework.

Security Assurance

Encryption technology and relevant policies need to be proposed to make data curation secure.

Value-added Information

For data curation, providing value-added information based on proper preservation is the desirable situation which is helpful to further research on relevant topics.

Conclusions

Since data curation is an emerging field for academic libraries, its future directions need to be studied to better satisfy the information needs of researchers. Based on several current measures, five aspects of directions are proposed, including collaboration and integration, permanent preservation, data quality framework, security assurance and value-added information.

Acknowledgments

Many thanks to the authors whose studies give great enlightenment to me, and the logo from UCSD Library. Thanks to the instructors of LIS501 Course: Maria Bonn, David Dubin and Emily Knox. Also, thanks for the help of Rebecca Hodson and the support of GSLIS.

ILLINOIS