Exploring Research Data Management from a Data User's Perspective

Yejun Wu[1][0000-0001-8411-4926]

[1] School of Library and Information Science
Louisiana State University, Baton Rouge, LA 70808, USA
wuyj@lsu.edu

Abstract. Current research data management practices focus more on data collecting, curating and sharing than on supporting data use and reuse. This research studies research data management from a data user's perspective and aims to reveal what research data service features may support data reuse. Two data repositories – the TREC website and the GRI data repository – were studied and compared from the perspectives of three types of data users (i.e., insiders, community users, and public users). The TREC website can support data reuse for insiders and community users, but not necessarily public users. The GRI repository can support data reuse for some insiders only. The findings have multiple implications for research data management.

Keywords: Research Data Management, Research Data, Data Repository, User

1 Introduction

The purpose of managing and sharing research data is to encourage its reuse. The goal of this research is to study whether current research data management (RDM) or research data services (RDS) provided by current research repositories support data reuse from a user's perspective, and study what RDS features may support data reuse.

2 Novelty and Significance of the Study

A literature review of current RDM research finds that most of the literature discusses RDM from service provider's perspective (on collecting, describing and curating research data) or from data contributors' perspective (on faculty researchers' current practices in describing, organizing, and sharing data and their needs for RDS and education). Faniel's team [3] studied data sharing and reuse behavior of social scientists, archaeologist and zoologists. No paper has been found to report RDM directly from data user's perspective. In addition, studying RDM from a user's perspective may reveal problems of RDM practices in supporting data reuse, and the findings of this study may be used to improve RDM to serve its purpose.

3 Objective and Method

The objective of this research is to study and compare a non-typical data repository that primarily focuses on data use and reuse to support a sub-disciplinary research task with a typical data repository that primarily focuses on the curation and sharing of cross-disciplinary research data. The purpose of the comparison is to identify which repository supports data use/reuse and what RDM features support data use/reuse, and provide insights into practicing RDM for data use/reuse purpose.

To fulfill the objective, I selected the National Institute for Science and Technology (NIST) Text Retrieval Conference (TREC) website (https://trec.nist.gov/) and the Gulf of Mexico Research Initiative (GRI) data repository (https://data.gulfresearchinitiative.org/data-discovery). The TREC website hosts the task guidelines, data, past results, scripts, publications, and FAQs of information retrieval system evaluation (sub-disciplinary) tasks. The two data repositories are studied according to Mayernik's [4] theoretical framework of five categories of "institutional carriers" for data practices: (a) norms and symbols, (b) intermediaries, (c) routines, (d) standards, and (e) material objects.

4 Preliminary Findings and Implications

In terms of the TREC website, the guidelines, publications and tools accompanying the datasets may have provided sufficient support to expected users to understand and use the datasets even though no metadata standards were applied to describe the datasets. The TREC website has established the five institutional carriers to support data reuse for insiders and community users, but not necessarily public users. Although the GRI data repository encourages data reuse and interdisciplinary research, it probably has not established the five institutional carriers to ensure data reuse by community users. It can support data reuse for some insiders only, but not necessarily community users and public users. The findings have implications for RDM in terms of setting a goal of providing what level of support for what type of users, designing different RDS for different types of data repositories (such as sub-disciplinary repository vs. cross-disciplinary repository), and probably providing extra services (such as education) to support public users.

References

1. Faniel, I., Yakel, E. 2017. Practices Do Not Make Perfect: Disciplinary Data Sharing and Reuse Practices and Their Implications for Repository Data Curation. In Johnston, L.R. (Ed.), Curating Research Data Volume 1: Practical Strategies for Your Digital Repository, 103-126 (2017). Chicago: Association of College and Research Libraries Press.
2. Mayernik, M. Research data and metadata curation as institutional issues. JASIS&T 67(4): 973-993 (2016).