Functionality Analysis of an Open Source Repository System: Current Practices and Implications

Hsin-liang Chen [a,*], Yin Zhang [b]

[a] *Department of Library and Information Science, College of Informatics and Computing, Indiana University-Indianapolis, USA*
[b] *School of Library and Information Science, Kent State University, USA*

ARTICLE INFO

ABSTRACT

The aim of this research study is to examine the functionality development of the open source repository system: DSpace. The data on DSpace repositories' implementation practices were collected from the DSpace User Registry during September 2013–March 2014. A total of 545 repositories in the registry indicated specific system function customizations, representing 533 unique institutions from 95 countries worldwide. The findings indicate that U.S.A. and India are the top two countries to have adopted DSpace. The majority of the DSpace digital repositories are created by academic institutions, which indicates a strong representation of academic institutions in the use of DSpace. The major adopted system functions are statistics, Dublin Core Meta Toolkit, Manakin Themes, and language packages. Most DSpace members use the repository system as their institutional and learning resource repositories. The top content types are conference papers, research documents, and learning/teaching materials. The implications of the findings are also discussed.

INTRODUCTION

Since the beginning of the 21st century, several Digital Library (DL) software systems have been built to serve different organizations. And as these DL systems have come to be widely implemented by different types of organizations to manage their information and assets, the systems have also come to serve different roles to meet the needs of these diverse organizations. DL systems serve as digital archives, digital museums or institutional repositories depending on how the systems are implemented by organizations.

Some DL systems such as DSpace, EPrints, Fedora, and Greenstone are free and open source systems that have their own member consortia worldwide. In order to reflect the diversity of system members, the term repository is used to include non-library members within DL systems. With this "open" approach, unique repository system functions have been created and shared among members. Van Westrienen and Lynch (2005) conducted a survey in 13 western countries and pointed out that repository systems were important components of campus infrastructure within the academic sector. The 2009 April/May issue of the *Association for Information Science Bulletin* devoted a special section on several issues regarding repository systems and academia (ASIST Bulletin, 2009). These issues included the roles of repository systems on campus, the managing and leading authority of repository systems,

the "open" structure of repository systems, and policies on repository systems.

After a decade of development of repository system functions, it is important to understand to what extent the created repository system functions have been adopted by members and what factors have influenced the adoption of the functions. Such understanding will help institutions to better implement a repository based on common practices with a similar purpose.

The aim of this project is to examine the functionality development of one open source repository system: DSpace. The decision to examine DSpace is because it has the largest repository user community and developers worldwide; it is completely customizable to meet the needs of individual institutions and repositories; and it was initially developed by academic institutions and now is most commonly used by research libraries as an institutional repository (DSpace, 2014). What makes this research project possible is the availability of DSpace's user registry on the website "dspace.org," where DSpace members report information about themselves, including organization status, collections, use cases, adopted system functions, and locations.

LITERATURE REVIEW

DEVELOPMENT OF DSPACE

Barton and Walker (2003) reported that DSpace was initially MIT's repository system which managed, hosted, preserved, and enabled distribution of the scholarly output of MIT's faculty. The DSpace project was a joint venture between the MIT Libraries and Hewlett-Packard (HP),

* Corresponding author at: Department of Library and Information Science, Indiana University School of Informatics and Computing, 535 W. Michigan Street, IT 475, Indianapolis, IN 46202.
E-mail addresses: chenhsin@iupui.edu (H. Chen), yzhang4@kent.edu (Y. Zhang).

made possible with a grant from the Andrew W. Mellon Foundation. Tansley et al. (2003) summarized DSpace's initial functions as a data model, metadata, e-people, authorization, ingesting, workflow, CNRI Handle system, search and browsing, Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), subscription, and Web user interface.

From its initial success, DSpace has grown into a worldwide community. DSpace members have added more functions to the system based on the initial system functions. For example, the Texas Digital Library team introduced Manakin for specialized user interfaces (Phillips, Green, Maslov, Mikeal, & Leggett, 2007), added a customized workflow management system and Open Archives Initiative Object Reuse and Exchange (OAI-ORE) (Mikeal et al., 2009; Maslov et al., 2010; Lagoze, Van de Sompel, Nelson, Warner, & Johnston, 2012), and created a Web 2.0-based interface for a map collection (Maslov, Mikeal, Weimer, & Leggett, 2009) to the DSpace system.

Semantics is another emerging development area in DSpace functions that aims to facilitate more efficient search processes among DSpace members and their collections (Cherukodan, Kumar, & Kabir, 2013; Kruk & McDaniel, 2009; Usman & Khan, 2012). Additionally, Cherukodan et al. (2013) applied Google Analytics to evaluate the distribution of the digital items and usage of an academic DL implemented by DSpace.

DSpace is well-documented in a Wiki site made up of contributions by its community (https://wiki.duraspace.org/display/DSPACE/Home). The Wiki site contains DSpace related resources, announcements, user community information, areas for developers, and official DSpace documentation with details about DSpace release notes concerning different versions, installation, administration, and system functions. These resources benefit developers, researchers, and users alike.

With emerging repository functions supporting research, learning and teaching activities in academic institutions, many academic librarians have envisioned and implemented changes at academic libraries as well as advocated new roles of academic libraries. After summarizing the development of repositories in higher education in the early 21st century, Walters (2007) pointed out that academic libraries were involved in managing diverse digital information resources ranging from external commercial databases to internal teaching and learning materials produced by faculty and students. Nagra (2012) offered a set of best practice guidelines for academic librarians to implement effective and successful institutional repositories on campus.

ADOPTION THEORY AND REPOSITORY SYSTEM DEVELOPMENT

Repository systems are implemented by various institutions as an innovative way to manage institutional properties more efficiently and effectively (i.e., Lynch, 2003; Westell, 2006). In order to facilitate the development of these innovation systems, Venkatesh and Davis (2000) reaffirmed that perceived usefulness and usage intention are key factors influencing users' system adoption behaviors. Furthermore, Munir and Kay (2003) identified that organizational culture and actual usefulness were two key factors in clinical information system integration.

In response to the user adoption issues, various approaches have been proposed by researchers and software developers in the course of system development. For example, Toleman, Ally, and Darroch, (2004) used adoption theory as their project base and applied agile system development methodologies to a new publishing system. Bergek, Jacobsson, Carlsson, Lindmark, and Rickne (2008) proposed a framework of analyzing the functional dynamics of technological innovation systems. This framework emphasizes the dynamics of an innovation system's functions, which directly influence the development of the innovation system.

GLOBAL IMPLEMENTATION OF REPOSITORY SYSTEMS

The implementation of repository systems has a great impact on the traditional scholarly communication system. Cullen and Chawner (2011) surveyed eight universities in New Zealand regarding the implementation of repository systems and reported different views on the traditional scholarly communication system and repository systems between the university faculties and academic libraries. They suggested that academic libraries should maximize repository systems' accessibility and citability in their institutional rewards systems.

Bhardwaj (2014) analyzed 436 repository research papers published between 2001 and January, 2013 and found that "developing countries are lagging behind not only in building repositories but also in publishing research about them," (p. 198). Since 2013, there have been publications from several developing countries. For example, DSpace was used at two institutions in India: Kuvempu University (Biradar & Banateppanavar, 2013) and Raman Research Institute (Meera, Manjunath, & Kaddipujar, 2013); Mapulanga (2013) described projects at the University of Malawi Libraries; Palmer, Bollini, Mornati, and Mennielli (2014) reported the DSpace-CRIS project at the University of Hong Kong; and Rahman and Mezbah-ul-Islam (2014) reported the recent repository development at universities in Bangladesh. A recent survey on Asian repository systems done by Loan (2014) indicated that "Asia is the third largest region in terms of number of open access repositories in the world after Europe and North America," (p. 35). Among the Asian countries, Japan is the number 1 contributor followed by India and Taiwan (Loan, 2014). He also pointed out that DSpace was the most adopted repository system (almost 69%) in the survey.

In summary, DSpace is the repository system most widely adopted worldwide. Over the years, various system functions have been developed to facilitate local institutions in enhancing the scholarly communication systems. Most studies reported on the adoption of various repository systems, repository research authorship, and the types of institutions and repository collections. There is a lack of research on repository system functions and their support to various collection types. As DSpace is the most popular repository system worldwide, the purpose of this project aims to study the adoption of its functions by global institutions in terms of various repository collections and use cases.

RESEARCH QUESTIONS

To form an understanding about the current practices of creating repository systems in general, and DSpace repositories in particular, the following two research questions are explored:

- RQ1: *What are the most adopted system functions for digital repositories?*
- RQ2: *How may system function adoptions vary by institution type, use case, and content type?*

METHODOLOGY

The data about DSpace repository implementation practices were collected from the DSpace User Registry (http://registry.duraspace.org/registry/dspace) during September 2013–March 2014. A total of 545 repositories in the registry indicated specific system function customizations, representing 533 unique institutions from 95 countries worldwide, with the United States and India being the top two countries with the largest number of repositories utilizing the customization functions. Table 1 summarizes the most representative countries in the sample.

The following data elements were collected for each of the repositories and the data were processed using Microsoft Access and SPSS for analysis:

- institution affiliation,
- institution type,
- country,
- use case type(s),
- content type(s) in the repository,
- file type(s) in the repository, and,
- system implementation integrations/customizations.