

01 Nov 2014

## System Function Adoption of an Open Source Digital Repository System: A Global View

Hsin-Liang Chen

Missouri University of Science and Technology, chenhs@mst.edu

Yin Zhang

Follow this and additional works at: [https://scholarsmine.mst.edu/library\\_facwork](https://scholarsmine.mst.edu/library_facwork)

 Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Chen, H., & Zhang, Y. (2014). System Function Adoption of an Open Source Digital Repository System: A Global View. *Proceedings of the 77th ASIS&T Annual Meeting (2014, Seattle, WA)*, 51(1) John Wiley & Sons Inc..

The definitive version is available at <https://doi.org/10.1002/meet.2014.14505101077>

This Article - Conference proceedings is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Library and Learning Resources Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact [scholarsmine@mst.edu](mailto:scholarsmine@mst.edu).

# System Function Adoption of an Open Source Digital Repository System: A Global View

**Hsin-liang Chen**

Department of Library and Information Science  
Indiana University-Indianapolis  
chenhsin@iupui.edu

**Yin Zhang**

School of Library and Information Science  
Kent State University  
yzhang4@kent.edu

## ABSTRACT

This poster reports a project that examines the adoption of system functions of an open source digital repository, DSpace. It also identifies the factors which have influenced the functions' adoption. The data were collected from DSpace user registry from September 2013 to March 2014. A total of 545 repositories in the registry contained system function customizations, representing 533 unique institutions from 95 countries. The preliminary findings indicate that 10 of the 32 available system functions are adopted by over 10% of its members; the majority of repositories are from academia; academic repositories also offer most system functions; and the U.S. and India each comprises over 10% of DSpace repositories that have DSpace system function customizations. Additionally, repositories from India utilize most system functions. About two-thirds of institutions are using DSpace as their institutional repositories and the two major (over 50%) content types are conference publications and technical reports.

## Keywords

Data management, digital repositories, systems analysis.

## INTRODUCTION

As digital library (DL) systems have been widely implemented by different types of organizations to manage their information and assets, those systems often serve different roles to meet the needs of those diverse organizations. For example, DL systems can be seen as digital archives, digital museums or institutional repositories depending on how those systems are implemented by those organizations. Since the beginning of the 21<sup>st</sup> century, several DL software systems have been built to serve different organizations. Some of those 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 "institutional repository" (IR) is used to include non-library members within the category of DL systems. With this "open" approach, unique IR system functions have been created and shared among members.

After a decade of development of IR system functions, it is important to understand to what extent the created IR system functions have been adopted by members and what factors have influenced the adoptions. The aim of this project is to examine the functionality development of one open source IR system, DSpace. The decision to examine DSpace is based on the availability of information about its members on the website "dspace.org," where DSpace members report information about repositories using the system, including organization status, country, collections, use case, contents, and adopted system functions.

## LITERATURE REVIEW

Tansley et al. (2003) summarized DSpace's 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 organization. Based on those initial system functions, more functions have been added to DSpace by its members. For example, the Texas Digital Library team introduced Manakin for specialized user interfaces (Philips, 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), and created a Web 2.0-based interface for a map collection (Maslov, Mikeal, Weimer, & Leggett, 2009).

Semantics is one emerging development area in DSpace functions that aims to facilitate more efficient search processes among DSpace members and their collections (Kruk & McDaniel, 2009; Usman & Khan, 2012; Cherukodan, Kumar, & Kabir, 2013). Additionally, Cherukodan, Kumar, and Kabir (2013) applied Google

*ASIST 2014, November 1-4, 2014, Seattle, WA, USA.*

©2014 Hsin-liang Chen and Yin Zhang

Analytics to evaluate the distribution of the digital items and usage of an academic DL implemented by DSpace.

At this stage in the development of IR systems, it is important to examine the adoption of created IR systems' functions and to identify the factors which have influenced the functions' adoption. Such research is useful for institutions, IR managers, system developers, and academic librarians, that plan to implement their DSpace repositories with the best practice strategies in mind.

### RESEARCH QUESTIONS

To form an understanding about the current practices of creating IR systems, the following questions are explored:

- RQ1: What are the most adopted system functions by DSpace member institutions?
- RQ2: How may system function adoptions vary by factors such as the institution type, country, use case, content type, etc.?

### RESEARCH METHODS

The data about DSpace repositories implementation practices were collected from the DSpace User Registry (<http://registry.duraspace.org/registry/dspace>) from September 2013 to March 2014. A total of 545 repositories in the registry indicated specific system function customizations, representing 533 unique institutions from 95 countries worldwide. For each of the repositories, the following data elements were collected:

- 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.

### RESULTS

The data analysis is still ongoing and expected to be completed in July 2014. The preliminary results are summarized in this proposal. Also, due to space limit of the proposal, only a select set of the results are included. Full and final results will be presented at the conference.

*RQ1:* What are the most adopted system functions by DSpace member institutions?

A total of 32 unique DSpace system functions are reported under integrations/customizations by the DSpace registry members. Table 1 summarizes the top 10 adopted functions. The most adopted function is Statistics (43%), which can make repository usage data available to administrator and repository visitors. The next three most popular functions are Dublin core Meta Toolkit (28%), Manakin Themes (27%), and Language Packages (23%). The other popular

functions that make the top ten are adopted by at least 12% of the repositories in registry.

Integration/Customization Functions	Number of repositories	Percent
1. Statistics	236	43%
2. Dublin Core Meta Toolkit	154	28%
3. Manakin Themes	147	27%
4. Language Packages	127	23%
5. Embargo	92	17%
6. Creative Commons Open URL	85	16%
7. DSpace Discovery	83	15%
8. Mirage	77	14%
9. SWORD	73	13%
10. Controlled Vocabulary-Ontology	64	12%

**Table 1. Top 10 most adopted system functions.**

*RQ2:* How may system function adoptions vary by factors such as the institution type, country, use case, content type, etc.?

#### *Customization functions by institution type*

As shown in Table 2, an overwhelming majority of the DSpace digital repositories are created by academic institutions (70%), followed by government (7%), research center (5%), and nonprofit (5%).

Repositories created by various institution types have utilized the 32 DSpace system functions quite differently. Table 3 shows the different range of customizations utilized by institution type. It appears that academic institutions have used all customization functions, while others have used only part of what are offered. While there are only few personal repositories, they tend to use a wider range of DSpace functions compared to other large number of repositories created by government, research center, and nonprofit.

Institution type	Number of repositories	Percent
Academic	380	70%
Government	40	7%
Research center	28	5%
Nonprofit	26	5%
Personal	16	3%
Archive/public library	13	2%
Commercial	11	2%

**Table 2. Most representative repositories by institution type.**

Institution type	Variety of customizations	Coverage percent
Academic	32	100%
Personal	29	91%
Government	28	88%
Research center	28	88%
Nonprofit	19	59%
Commercial	17	53%
National library	15	47%
Archive/public library	14	44%
Medical center hospital	13	41%
Consulting / service provider	5	16%

**Table 3. Function adoption by institution type.**

#### *Functions by country*

The repository sample contains digital repositories from 95 countries, with the United States and India being the top two countries with the largest number of repositories utilizing the customization functions (see Table 4).

As shown in Table 5, there is a range of variety of customization functions utilized by country, with Indian repositories utilizing almost all available functions. Interestingly, while the United States has the largest number of repositories in the sample, its repositories have adopted only 25 out of 32 (78%) functions.

Country	Number of repositories	Percent
1. U.S.A.	68	12%
2. India	57	10%
3. Spain	28	5%
4. United Kingdom	21	4%
5. Brazil	19	3%
6. Colombia	14	3%
7. Taiwan	14	3%
8. Vietnam	13	2%
9. Ukraine	12	2%
10. Indonesia	12	2%
10. Canada	12	2%

**Table 4. Most representative repositories by country.**

Country	Variety of customizations	Coverage percent
India	31	97%
United Kingdom	29	91%
Indonesia	27	84%
Vietnam	26	81%

Brazil	25	78%
U.S.A.	25	78%
Mexico	24	75%
Philippines	24	75%
Colombia	21	66%

**Table 5. Function adoption by country.**

#### *Functions by use case*

There are nine different use cases (see Table 6) for DSpace digital collections, among which institutional repository is the most used type (68%), followed by learning resources (30%), image repository (23%), and subject repository (21%). Some repositories belong to more than one use case category.

Among the 32 DSpace customization functions, repositories in different use cases seem to all utilize these functions widely (See Table 7). A close look at the most adopted customization functions are similar across all use cases.

Use case	Number of repositories	Percent
1. Institutional Repository	373	68%
2. Learning Resources	161	30%
3. Image Repository	123	23%
4. Subject Repository	115	21%
5. Audio/Video Repository	106	19%
6. Museum/Cultural Heritage	63	12%
7. Government Records/Reports	60	11%
8. Other	40	7%
9. Federated Repositories/Networked Instances	30	6%

**Table 6. Repositories by use case.**

Use case	Variety of customizations	Percent
Learning Resources	32	100%
Subject Repository	32	100%
Audio/Video Repository	31	97%
Image Repository	31	97%
Institutional Repository	31	97%
Other	31	97%
Museum/Cultural Heritage	30	94%
Federated Repositories/Networked Instances	29	91%
Government Records/Reports	29	91%

**Table 7. Function adoption by use case.**

### Functions by content type

There are 15 different content types for the DSpace digital repository registry, eight of which have over 100. Some repositories are in more than one content type. Conference papers/presentations and technical reports/work papers appear to be the two most common repositories at this point, each representing more than half of the repositories in the registry.

When examining the variety of customization functions adopted by content type, interestingly, all types of content utilize at least 30 out of 32 functions, suggesting content type does not contribute to different system customizations.

Top repositories by content type	Count	Percent
1. Conference papers and presentations	290	53%
2. Technical reports / work papers	277	51%
3. Learning objects & resources	237	43%
4. Subject/Special Collections	192	35%
5. Research or development organization project	178	33%
6. Reference Documents	139	26%
7. History/Art Archives	136	25%
8. Data sets	115	21%

**Table 8. Most representative repositories by content type.**

### DISCUSSION

In this study, we have identified the top ten system functions adopted by DSpace members. Four of those ten functions are used by over 20% of the members. As statistics is the most popular function used by over 40% of members, we can assume that the institutions of the repositories want to track all kinds of repository activities and to assess the performance of their repositories.

We have also learned that Dublin Core Toolkit and Manakin Themes are used by over a quarter of the members, which makes Dublin Core the most popular metadata standard among DSpace members. The adoption of Manakin Themes indicates that those members want to create more search interfaces to assist users. As DSpace members are from different countries, it is not surprising to see that language packages as the fourth most popular function.

In terms of membership, academic institutions from the U.S. and India are the major players in the DSpace community. Indian institutions tend to adopt more system functions compared to their U.S. counterparts.

DSpace has been used by most members as their institutional repositories storing the institutions' conference publications and technical reports in terms of content type. Additionally, DSpace has been applied in various use cases and content types with an almost equal wide range of adopted system functions. Repositories in different content types and for various use cases appear to adopt an almost equally wide range of DSpace system functions.

### CONCLUSION

So far, we have examined the adopted system customization functions and the range and extent of such adoption among the DSpace digital repositories. We have also identified possible factors that contribute to the differences in adopted system functions: Institution type and country. Further analysis will be conducted to examine the specific customizations adopted among the digital repositories and identify possible factors for different customizations.

### REFERENCES

- Cherukodan, S., Kumar, G. S., & Kabir, S. H. (2013). Using open source software for digital libraries: A case study of CUSAT. *The Electronic Library*, 31(2), 217-25. doi:10.1108/02640471311312393
- Kruk, S. R., & McDaniel, B. (Eds.). (2009). *Semantic digital libraries*. Berlin: Springer.
- Maslov, A., Creel, J., Mikeal, A., Phillips, S., Leggett, J., & McFarland, M. (2010). Adding OAI-ORE support to repository platforms. *Journal of Digital Information*, 11(1).
- Mikeal, A., Creel, J., Maslov, A., Phillips, S., Leggett, J., & McFarland, M. (2009, June). Large-scale ETD repositories: A case study of a digital library application. In Heath, F., & Rice-Lively, M. L. (Eds.). *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries: June 12-15, 2009, Austin, Texas, USA*. (pp. 135-144). New York: ACM. doi:10.1145/1555400.1555423
- Phillips, S., Green, C., Maslov, A., Mikeal, A., & Leggett, J. (2007). A New Face for DSpace. *D-Lib Magazine*, 13(11/12), 1082-9873.
- Tansley, R., Bass, M., Stuve, D., Branschofsky, M., Chudnov, D., McClellan, G., & Smith, M. (2003, May). The DSpace institutional digital repository system: current functionality. In *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries* (pp. 87-97). IEEE Computer Society.
- Usman, A., & Khan, S. (2012). Exploiting semantics in subject-based searching of DSpace. In Chen, H.-H., & Chowdhury, G. (Eds.), *The Outreach of Digital Libraries: A Globalized Resource Network*: