Introduction to the Special Issue of "Digital Collection Metadata & Internet Discovery"

Le Yang & Joy M. Perrin

Digital repositories and digital collections have increased in size tremendously in recent decades in academic and research libraries. As digital items are put into digital repositories in order to be discovered by search engines, associated metadata records need to be effectively set up for external indexing. Some discussion has begun in the existing literature about digital resources discovery, metadata evaluation, search engine indexing, and search engine optimization strategies. However, due to the distinct options of digital repository software, the complexity of metadata schemas, the variety of formats of digital items, and the ambiguity of search engine indexing strategies, researchers have not yet reached an agreement about which metadata schema is the best strategy to use, for which formats of digital files, in which repository system, or for which search engines. This special issue aims to explore these approaches and offer insight into the current literature debating digital collection metadata and its discoverability on the Internet, thereby offering a platform for researchers to discuss metadata strategies, digital repositories, digital formats, search engine indexing, and Internet discovery.

Information seeking behaviors have been greatly shaped by Internet search engine dominance. As previous research has shown, metadata implementation strategies were inevitably developed based on rules of search engine optimization (Arlitsch and O'Brien 2012; Beel et al., 2009; Coates 2014; Yang 2016). One of the most important duties for digital librarians is to evaluate indexing effects of search engines on their local digital resources. By doing so, digital librarians are able to adjust metadata strategies to improve the discoverability of digital content on the Internet.

One such approach is discussed in this special issue. In a statistical study of digital repositories at a national level, Joel Alhuay-Quispe and team (2017) from the Universidad San Ignacio de Loyola analyzed the ALICIA indexing ratio and the Google Scholar indexing ratio for 48 institutional repositories' digital items in Peru, including 10 public universities and 38 private ones. ALICIA is Peru's national digital repository of science, technology, and innovation, which was designed to harvest digital resources from domestic universities' IRs. The research finds that ALICIA does not have different indexing preferences based on type of university or format. In contrast, Google Scholar indexes more digital items from private universities than public ones, though both types of university IRs are not indexed well by Google Scholar. The authors

suggest that restriction of access and limitations of computer capacity might be the reasons. Electronic theses are well indexed in ALICIA, but document types such as articles, books, and conference proceedings are poorly harvested in ALICIA, because metadata standard DRIVER Guidelines and OAI-PMH are not followed strictly.

In contrast to ALICIA's harvesting mechanism from local digital repositories, Digital Public Library of America (DPLA) implements a more complex and structured model that consists of a content hub and a service hub. Each metadata aggregator service hub functions similarly to ALICIA and feeds highly aggregated metadata to DPLA, which inevitably causes issues of multi-layered metadata records. Krystyna Matusiak (2017) from the University of Denver conducted a user study on information seeking behavior and the navigation discovery system through the DPLA interface. The study finds that, although most study participants could navigate the DPLA's distributed multi-layered system effectively, some of them still encountered some confusion when navigating through a three-step process to locate digital objects provided by a metadata aggregator service hub. The author suggests that consistency in user interfaces of service hubs can improve user experience with distributed large-scale digital library systems.

Updating metadata strategies improves discoverability of digital collections; however, metadata is the only one of many key factors that affects indexing and discoverability of digital resources on the Internet. The digital repository system is another key factor that impacts Internet discovery of digital resources. At a digital conference in 2013, Perrin explained why her institution migrated digital content from one system to another, which resulted in increased visit counts through organic search engines. Similarly, in this special issue a team of library science scholars from the University of Utah, led by Anna Neatour (2017), introduces a digital asset management system migration project from CONTENTdm to homegrown Solphal. The article shares details on reviewing and selecting tools for metadata implementation and digital asset management. In the migration project, the authors provide insightful strategies on large-scale metadata cleanup, standardization, management, and discovery enhancement. The success of system and metadata migration enables a streamlined workflow and reduces human errors for collection managers; it also improves user experience by reducing indexing response time, easing site navigation, and enhancing discovery of digital resources.

It is convenient for digital librarians to make use of an existing metadata schema for their digital collections. However, what if no metadata schema has previously been created for a specific type of digital resource? One unique project in this issue discusses Oregon State University's historical clothing collections. Maura Valentino (2017) notes that clothing collections have been serving as important resources for researching and teaching in social science fields, such as

fashion design and sociology; however, no metadata schemas were ever created for digital clothing collections. In order to address this concern, Valentino created a linked data metadata schema for clothing designed for a Historical and Cultural Textile and Apparel Collection that is used as a teaching and research resource in Oregon State University's Design and the Human Environment department. The author also proposes that other organizations can implement this metadata scheme for their digital clothing collections.

Digital archives feature one of the most important applications of digital scholarship. Exemplar case studies and best strategies guide peer librarians and archivists in developing digital collections. Ian Goodale (2017), the European Studies and Digital Scholarship Librarian at the University of Texas at Austin, introduces an interesting digital scholarship project, detailing the development, promotion, and future improvement of the Prague Spring Archive. The completion of this digital project provides researchers open, online access to historical documents related to one of the key moments in the Cold War – the Prague Spring. To address online discoverability, the author proposed to implement full-text XML based on OCRed content in its selected Scalar digital system.

Another featured digital archive project in this issue is Cillian Joy and collaborators' project (2017). The article describes how archivists at the National University of Ireland, Galway developed an archival project on Tim Robinson, a cartographer and writer who studied the landscape surrounding Galway Bay for over 40 years. The authors provide details on technical specifications and selection of a preservation tool, discovery layers, linking and digital mapping infrastructure, and implementation of different types of metadata. The digital interface, which allows users to interact and explore on a digital map, enhances user experience in teaching, learning, and researching.

In addition to digital resources, academic libraries spend millions of dollars for subscriptions to electronic resources. However, electronic resources in subscribed databases are usually not organic-searching friendly. From an organic search platform, library patrons may never know what resources were already made available for them by their academic libraries. In order to address this issue, Jason Clark and Doralyn Rossmann (2017) described a successful metadata optimization project at Montana State University Library. Through the Open SESMO strategy, they applied search engine optimization and structured data, linked data models, and social media optimization techniques to all their subscribed databases. This innovative practice shows a significant return-on-investment with increased traffic to databases from organic search referrals.

This special issue of “Digital Collection Metadata and Internet Discovery” is only the beginning of the journey exploring discoverability issues of valuable digital resources on the Internet. As digital librarians working on digital repositories, the guest editors hope that professionals in the field can benefit from these articles and thus set priorities for local digital collection’s development strategies. However, more work remains to be done – specifically more research is needed to better understand concerns observed by digital repository librarians in their day-to-day work. For example, the editors still hope to see evaluation and comparison studies of different repository systems, which could help librarians make decisions based on the technological options available. Moreover, there are no indexing studies on different formats and varied types of digital resources that can help peers develop better digital collections for specific usage. It is hoped that future research and case studies can be added to enrich the literature of the field.

References

Arlitsch, Kenning, and Patrick S. O’Brien. 2012. “Invisible Institutional Repositories: Addressing the Low Indexing Ratios of IRs in Google Scholar.” *Library Hi Tech* 30 (1):60–81. doi:10.1108/07378831211213210.

Beel, Jöran, Bela Gipp, and Erik Wilde. 2009. “Academic Search Engine Optimization (ASEO) Optimizing Scholarly Literature for Google Scholar and Co.” *Journal of Scholarly Publishing* 41 (2):176–90. doi:10.3138/jsp.41.2.176.

Coates, Mildred. 2014. “Search Engine Queries Used to Locate Electronic Theses and Dissertations: Differences for Local and Non-Local Users.” *Library Hi Tech* 32 (4):667–86. doi:10.1108/LHT-02-2014-0022.

Perrin, Joy. 2013. CONTENTdm to DSpace - Why?. Texas Conference on Digital Libraries.

Yang, Le. 2016. “Making Search Engines Notice: An Exploratory Study on Discoverability of DSpace Metadata and PDF Files.” *Journal of Web Librarianship 10* (3):147-60. doi: 10.1080/19322909.2016.1172539.