Three Rs of Digital Collections

Catherine E. Hall, Robin A. Naughton, Xia Lin
The iSchool at Drexel
College of Information Science and Technology
Drexel University
Philadelphia, PA 19104 USA
+1.215.895.2482
{chall,rnaughton,xlin}@ischool.drexel.edu

ABSTRACT
People have a difficult time finding relevant digital collections even though there has been significant increase in openly accessible digital collections. In this poster, we describe our 3R system design for a digital collection repository that will facilitate user identification and interaction with digital collections, including mechanisms for *reviewing*, *ranking*, and *recommending* collections for the benefit of a social community.

Categories and Subject Descriptors
H.3.7 **[Information Storage and Retrieval]**: Digital Libraries – *collection, dissemination, standards, systems issues, user issues*.

General Terms
Management, Performance, Design, Reliability, Human Factors

Keywords
Digital libraries, recommender system, information discovery

1. INTRODUCTION

In recent years there has been an explosive increase in digital collections. Independent and federally funded initiatives have helped public and private institutions make available digital content of cultural, educational and historical significance: libraries and associations have established collections for their patrons [3][4], and agencies such as the Institute of Museum and Library Services (IMLS) have administered grants for digital collection creation and development [9]. In a parallel development, a number of digital material discovery tools such as OAIster [6] and OpenDOAR [7] have emerged to help people find what they need.

Despite the prevalence of digital collections, it is difficult for people to find collections which match their specific information needs. When people rely on search engines to find collections, the most relevant and comprehensive resources are not always returned by the search engines. This, in part, is due to the lack of agreed upon standards for developing and adopting metadata schemas for finding digital collections as well as individual items. In this poster, we report our ongoing research on the evaluation of digital collections and the development of collection finding tools. At the collection level, users need help from the social community to identify, evaluate, and understand the purpose, scope, and quality of each collection. We are designing a digital collection repository system that will facilitate user identification and interaction with digital collections, including mechanisms for *reviewing*, *ranking*, and *recommending* collections for the benefit of others. These three Rs are the key to finding and maximizing the usage of digital collections.

2. CHALLENGES OF FINDING COLLECTIONS

Digital collections provide a window to a wealth of previously restricted collections. With digital materials, people are no longer bounded by time or geography, giving scholars and students access to more quality material. However, research has shown that students are often unaware of databases within their fields until told about them [1], so we can expect that people are similarly unaware of excellent digital collections. Libraries and associations advertise their own collections but rarely have the time or resources to identify and evaluate others. Searching for collections on the Web is also problematic; there is no dominant standard for collection level metadata [5] and many content providers do not use any at all, making retrieval by search engines difficult.

Several tools have been developed in recent years to help users locate digital materials. OAIster [6] is perhaps the largest and best known tool, but its drawback is harvesting resources using item, not collection, level metadata, making a collection search

frustrating. Using OAIster, we searched for five digital collections (two each from Harvard and West Virginia Universities, and one from Amherst College); we found individual items from three of the collections, but the item records gave no indication that these pieces were part of larger collections which may also be useful to the user. Two of the collections - University of Virginia's Frances Benjamin Johnston Collection and Amherst College's Jerry Cohen AC 1963 Papers - could not be found using OAIster, even when searching at the item, rather than collection, level.

OAIster harvests item records using OAI-PMH, but not all creators of digital collections have adopted this NISO endorsed protocol; in 2007, only 19% of IMLS National Leadership Grant funded digital collection projects were using it [10]. Where OAIster identifies digital items, several registries identify collections. However, these have been criticized by NISO who claim that most appear to be poorly maintained [5]. Digital Collections and Content [9] is one such registry. The collections contained are easily found and well described, but are limited to those developed with assistance from IMLS funding or those with an emphasis on state history. As such, a wide variety of useful collections are missing. Similarly, OpenDOAR [7], a European initiative listing open access academic and subject repositories omits rich digital collections from many of the leading US universities and public institutions.

When looking for digital material, people's initial searching strategies are vague: they want to find a range of useful material on a subject rather than any particular item. OAIster offers a solution to finding items within a small and well-defined area of interest, but no solution for the user with an imprecise or broad interest. There is a clear need for a tool that identifies digital collections, and helps users to understand and evaluate them. Using the 3R's of reviewing, ranking, and recommending, our system will enable users to easily find resources that match their research needs.

3. DIGITAL COLLECTIONS EVALUATION CRITERIA

Finding collections is only a first step. Without mechanisms for assessing and evaluating collections, users cannot judge the collection value, whether the content meets their requirements, if the information can be trusted, or if the collection is easy to navigate. Search engines and existing aggregators of digital materials do not return results based on these criteria; instead, each user needs to invest a considerable amount of time and effort to explore each collection. Our system incorporates user submitted reviews of collections which will enable subsequent users to make meaningful evaluations. Evaluation of digital collections has occurred for as long as collections have existed, but traditionally they have been judged against criteria for physical libraries, human-computer interaction and information retrieval systems [12]. It is important that digital collections are measured against their own criteria, and that this is user-defined. To incorporate useful information in reviews we are establishing criteria for users to consider when writing reviews. As part of a Digital Libraries class at Drexel's iSchool, students were asked to evaluate digital collections and explain the criteria used to review them and why it was important. We then compared the results with criteria developed from similar studies by Xie [11] [12].

Table 1: Types of Evaluation Criteria

Collection Content Quality Quantity Currency Scope Authority Completeness	**Services** Help/access guides Links between related items Supplementary/supporting materials Other added value features
Usability Search/browse features Interface design Organization of materials Accessibility	**Management** Mission Statement Targeted Audience Contact information

Finding similarities between the evaluation criteria produced from our own study and those of Xie, we clustered results around four main categories: collection content, usability, services, and management (Table 1). Previous research, including the studies by Xie, included system performance as a separate criterion but Drexel students incorporated this within usability, a pattern we have followed here. The identified criteria can be used to underpin the reviewing, ranking and recommending processes.

4. DESIGN OF A DIGITAL COLLECTIONS 3R SYSTEM

We have been designing and testing a digital collections 3R system to help users review, rank, and recommend digital collections within a social community. Creation of a social community and use of recommender system technology will make our system a significant improvement over existing collection finding tools. A hybrid recommender system with both content-based (compares item descriptors for recommendation) and collaborative filtering (makes recommendations based on user information) methods will help accomplish the goals of the digital collections 3R system. It will build a social community that creates transparency and shared knowledge about digital collections similar to WikiLens, a community-maintained recommender system [2] and TechLens, a hybrid recommender system that recommends research papers [8].

4. 1 Accessing Digital Collections

For collection-based retrieval, it is important to let users specify criteria to search both the content of collections, and reviews, rankings and recommendations (3Rs). In our system, the search results will be presented together with collection ranking, collection reviews, and recommendation rates. The system also allows users to save searches to create personal collections. With these features, users can quickly access digital collections through searching, browsing, personal collections and personalized recommendations.

4. 2 Evaluating & Reviewing Digital Collections

People using existing collection finding tools do not know how collections were evaluated, how they compare to similar collections, or how they are viewed by the community. Our resolution allows users to review collections through systematic online evaluation forms covering content, services, usability and management criteria (see Table 1). For example, one user can evaluate the collection content, while another evaluates the usability of the same collection. Similarly accomplished are ratings where users can judge collections based on a scale of 1 (low) to 10 (high). From the ratings, the system calculates and automatically ranks or recommends collections.

4.3 Building Digital Collections Recommender Community

People make recommendations to each other regarding digital collections, but few current collection tools support this function. With our tool, users can form groups and networks, recommend collections, and receive system generated recommendations based on user preferences, user connections and collection descriptors. Similar to WikiLens' "Tell a Friend" and "Buddies" features, recommendations can be sent to both registered and non-registered users. By developing the digital collection 3R recommender system, we are developing a social community for users; an invaluable resource for finding and maximizing the use of digital collections.

5. ACKNOWLEDGEMENTS

Catherine Hall and Robin Naughton are supported by IMLS PhD Fellowships.

6. REFERENCES

[1]. Chu, C. S., and Law, N. (2005). Development of information search expertise: research students' knowledge of databases. Online Inform Review 29, 6 (2005), 621-642.

[2]. Frankowski, D., Lam, S. K., Sen, S., Harper, M. F., Yilek, S., Cassano, M., and Riedl, J. (2007). Recommenders everywhere: the wikilens community-maintained recommender system. In *WikiSym '07: Proceedings of the 2007 international symposium on Wikis*, 47-60.

[3]. Harvard University. (n.d.). A selection of web-accessible collections at Harvard University. (Web site). http://digitalcollections.harvard.edu/

[4]. Library of Congress. (n.d.). Digital collections and programs. (Web sites). http://www.loc.gov/library/libarch-digital.html

[5]. NISO Framework Working Group. 2007. A Framework of Guidance for Building Good Digital Collections. 3rd Edition. National Information Standards Organization. http://framework.niso.org

[6]. OAIster (n.d.). OAIster home page. http://www.oaister.org/

[7]. OpenDOAR (n.d.). OpenDOAR home page – Directory of Open access repositories. http://www.opendoar.org/

[8]. Torres, R., McNee, S. M., Abel, M., Konstan, J. A., and Riedl, J. 2004. Enhancing digital libraries with TechLens+. In Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries, pages 228-236.

[9]. University of Illinois at Urbana-Champaign (n.d.). IMLS Digital collections and content. (website). http://imlsdcc.grainger.uiuc.edu/

[10]. Urban, R. (2007). A Good Visual Resource Is Hard to Find. Presented at VRA 2007, "From Fair Use to Fair Trading: Creating a Digital Image Matchmaking Commons." Kansas City, Missouri, March 28, 2007. http://imlsdcc.grainger.uiuc.edu/docs/rjurban_VRA2007_matchmaking.pdf

[11]. Xie, H.I. (2006). Evaluation of digital libraries: Criteria and problems from users' perspectives. Library & Information Science Research, 28 (2006), 433-452.

[12]. Xie, H.I. (2008). Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment. Information Processing & Management, 44 (2008), 1346-1373.