

# Assessment of Digitized Library and Archives Materials:

## A Literature Review

Elizabeth Joan Kelly

J. Edgar & Louise S. Monroe Library

Loyola University New Orleans

### Abstract:

Professional literature about the assessment of digital libraries reflects a growing interest in both improving the user experience and in justifying the creation of digital collections to multiple stakeholders. This article explores some of the current key themes in digital library assessment literature through a review of current literature (2004-2014) gathered from both scholarly and popular resources online. The majority of scholarship about digital library assessment utilizes usability testing and Web statistics for data collection, while studies about altmetrics, the reuse of digital library materials, cost benefit analysis, and the holistic evaluation of digital libraries are also present in the literature. Exploring the literature about digital library assessment allows libraries to create effective and sustainable evaluation models based on the successes and short-comings of previously completed projects.

### Keywords:

digital libraries, digital libraries—evaluation, digital libraries--use studies, digitization of library materials, digitization of archival materials, usability, Web analytics

### Article Classification:

Literature Review

This is a post-print [post-peer review] version of a manuscript originally published in *The Journal of Web Librarianship* 8, no. 4 (2014). <http://www.tandfonline.com/doi/abs/10.1080/19322909.2014.95474>

# Assessment of Digitized Library and Archives Materials:

## A Literature Review

### Introduction

“If we build it, will they come?” is the oft-repeated question in the development of digital libraries. However, recent literature reflects that digital library stakeholders are not just interested in whether or not users are finding their materials but also in how and why, and in the quality of the user experience. A focus on assessment in higher education has intensified in recent years just as shrinking budgets have necessitated that libraries find ways to justify their existence. Showing that digital libraries are not only effective and appreciated by users but also cost efficient is becoming increasingly important. In order to develop a digital library assessment program, it is helpful to review different methods to evaluate digital libraries. This article will explore recent publications on the subject of digital library assessment with a focus on digitized library materials in order to aid digital library stakeholders in developing assessment plans based on the conclusions and best practices found in the professional literature.

Digital libraries (DLs) are often defined broadly to include many types of online resources such as e-journals, digitized cultural heritage materials, institutional repositories, and even library Web sites. These have their own challenges when it comes to assessing their success—or lack thereof—in connecting users to resources. For the purposes of this article, “digital library” will refer to digitized library and archival materials, or “multimedia digital

libraries” that are stored along with metadata in a database for purposes of information retrieval (Comeaux 2008, 461).

## **Methods**

The digital library literature was explored by conducting library database, catalog, and Web searches in portals including ABI/INFORM Global; Education Resources Information Center (ERIC); Google Scholar; and Library, Information Science & Technology Abstracts (LISTA). In order to represent current literature, only works published between 2004-2014 were reviewed. While this article does not seek to be comprehensive in its coverage, the literature reviewed has been included both for its informational value and for its ability to accurately represent the current scholarship. A 2009 bibliographic analysis of literature published about DLs between 1997-2007 reflects an overall upward trend in articles published about DLs, with usability, organizational and economic issues, and legal issues the most prominent subjects (Liew 2009, 248-251). A more recent article reviewed only digital collection studies published in 2012 and found that scholarship at that time focused on qualitative (survey responses, social tags) and quantitative data (Web analytics, use statistics, number of citations) (Todd-Diaz and O’Hare 2014, 257-258). This literature review revealed that the vast majority of DL assessment publications focused on usability and user studies (16 articles) and Web analytics studies (13 articles). The remainder of publications found on assessing DLs were on topics including altmetrics, reuse of DL materials, cost benefit analysis, and the holistic evaluation of DLs using several of the afore-mentioned methods. The current literature on each of these topics will be explored below.

## Usability and User Studies

Usability involves the formal testing of a product (or prototype of a product). Regarding DLs, usability testing is employed to find out how DLs are used in practice by actual users. Comeaux describes “user-centered design” as identifying target audiences (also known as a user study) and then conducting formal testing of the Web site (2008, 458-459). User information gathering is often mislabeled as usability; usability is actually “carefully observing users interacting with the system in a realistic way,” either in person, using screen-capture software, or both (Comeaux, 459). In “Evidence-Based Practice and Web Usability Assessment,” Frank Cervone noted that most libraries don’t have sustainable usability plans in place (2014, 11). He advocates for “evidence-based practice,” “... an approach to information practice that promotes the collection, interpretation, and integration of valid, important, and applicable user-reported, researcher observed, and research-derived evidence (Booth, 2001);” or using research instead of anecdotes or “common sense” (Cervone, 12). Evidence-based practice is an iterative process, or one in which testing is done, changes are made, and then the product is retested, forming a closed loop. A Web usability evidence-based lifecycle is made up of five looping stages: define the problem, find evidence, evaluate evidence, apply the results of the evaluation, and then evaluate the changes (Cervone, 12). Cervone also advocates for user-centered design as it is only through knowing how users expect to use a site and what previous experience and skills they possess that we may know what research questions to ask and, therefore, what problems to solve (Cervone, 14).

Many libraries may find themselves conducting usability studies of DLs. In order to not reinvent the wheel with each new study, several evaluation models have been proposed that

are applicable to a wide variety of DLs. The literature on usability in DLs remained small until 1999 (Jeng 2005, 99). In developing a model that examines effectiveness, efficiency, satisfaction, and learnability, Jeng defined a number of dimensions of DL usability, including interface effectiveness and design; usefulness, usability, and ease of use; system performance, system functions, user interface, reading materials, language translation, outreach program, customer customization options, installation, field maintenance, advertising, support-group users; inherent usability (functional, dynamic) versus apparent usability (visual impressions); and usability (product works quickly and easily) versus functionality (product does what it's supposed to do, but not necessarily well) versus accessibility (availability) (Jeng, 96-98, 101). A number of methods can be used to explore these dimensions including formal usability testing, usability inspection, card sorting, category membership expectation, focus groups, questionnaires, think aloud, analysis of site usage logs, cognitive walkthrough, heuristic evaluation, claims analysis, concept-based analysis of surface and structural misfits (CASSM), paper prototyping, and field study (Jeng, 99). Jeng's model was tested on two academic library Web sites and found the strength of relationship between effectiveness and how many steps were required to complete a task and between effectiveness and satisfaction were strong, and the strength of relationship between effectiveness and time to complete a task was medium to strong (Jeng, 99). While this model was tested on academic library Web sites, which do not fit the definition of DLs utilized in this article, Jeng's model is referred to frequently throughout the literature and has been used by DLs that do fit the criteria of this article and was therefore included.

Another proposed evaluation model identifies users' criteria for a successful DL and then applies them to the evaluation of existing digital libraries. This model was developed after finding that existing criteria for assessing DLs was largely based on traditional library evaluation criteria rather than specialized for DLs. In this model, Hong (Iris) Xie asked, "What criteria do users identify as important for the evaluation of digital libraries?" and "What are the problems with the existing digital libraries?" (2006, 434). Subjects were asked to identify essential criteria for the development and use of DLs, and then apply that criteria to the evaluation of existing DLs including but not limited to the Library of Congress (LOC) American Memory digital collection, the ACM (Association for Computing Machinery) Digital Library, and the SUNY-Buffalo Electronic Poetry Center (Xie, 438). The results showed that the developed criteria echoed those previously proposed in the literature as well as those used in other studies, but that the criteria showed a greater tendency towards the perspective of the user than of the developer (Xie, 446).

In 2008, Xie revisited the evaluation criteria by conducting a usability study employing a diary, questionnaire, and survey of two DLs: LOC American Memory and the University of Wisconsin Digital Collections, while also having the users rate the importance of the different facets of the previously developed criteria (1352). While the previous study found that users identified usability and collection quality as the most important facets of the evaluation, users in this second study rated usability and system performance as the most important qualities, possibly because the caliber of the DLs was high enough that the users trusted that the content was reliable (Xie, 1370).

The vast majority of the literature on DL usability focuses on case studies of individual institutions for purposes of refining their own DLs. An exception is the study of cultural usability, or how different cultures respond to the same DL. Howard Gardner's Multiple Intelligences (MI) theory, was tested as a framework for evaluating subjective cultural factors, but the author notes that this was less than successful (Smith 2006, 229, 237). Still, a framework for designing DLs with cultural differences in mind is identified as a necessity for guiding global digital library design (Smith, 229, 237).

Several articles on DL usability self-identify as using heuristic evaluation, in which users evaluate an interface and then judge it using specific criteria (heuristics). Some advantages to heuristic evaluation are that it is easy to conduct, minimal data analysis is required, it uncovers usability problems not uncovered in other usability forms, and the evaluators do not need to be specialists (Long, Lage, and Cronin 2005, 335). Staff at the University of Colorado Boulder employed Nielsen's ten usability heuristics to aid in planning an interface redesign for their aerial photographs digital collection based on principles of user-centered design (Long, Lage, and Cronin). In another study, both Nielsen's heuristics and ISO Heuristics 9241 were used to inspect the World Digital Library, Europeana, the British Library, Scran (Scotland), and the University of Edinburgh's instance of Aquabrowser to gauge the overall usability standard of established DLs, identify positive examples of good usability, recognize the unique aspects of each DL, and evaluate to what extent they enhance the user experience for the JISC-funded project Usability and Contemporary User Experience in Digital Libraries (UX2.0) (Paterson and Low 2010).

The remainder of recent literature on the usability of DLs focuses on a variety of different research problems. Maggie Dickson's 2008 study looks at the usability of one specific digital asset management system (DAM), CONTENTdm, to evaluate whether the platform meets user's needs, has an intuitive interface, and provides a satisfactory experience for users. After analyzing the search experiences of users and administering a follow-up questionnaire, it was determined that CONTENTdm's interface was confusing even for experienced online researchers and that promotion of digital collections, an issue not native to CONTENTdm but to the library Web site, was lacking (Dickson, 369). Users also experienced navigational issues within CONTENTdm that could be alleviated through enhanced item metadata and tutorials (Dickson, 370-371).

"Help" features of six digital libraries were explored in one study to determine the usability of just one aspect of a DL (Xie 2007). Help features were defined as "any features that assist users to effectively use digital libraries except general search and browse functions" (Xie, 879). Within the selected DLs (American Memory, New York Public Library Digital, International Children's Digital Library, Perseus Digital Library, American Museum of Natural History Digital Library Project, and Medline Plus), seven categories of Help features were identified: general, search-related, collection-related, navigational, terminology-related, customizable, and view-and-use related (Xie 865, 869), with four presentation styles: descriptive, guided, procedural, and exemplary (Xie, 873). Finally, Xie identified six common problems among DL Help features: lack of standards, tradeoff between using explicit Help (any feature with "Help" or "?" in the label) and implicit Help (features that assist users but are not labeled explicitly), tradeoff between using general Help (FAQs and Contact Us) versus specific Help (pertaining to a single

collection), lack of interactive Help features, lack of dynamic presentation styles, and lack of Help features for advanced, non-English speaking users (877).

Identifying the success of reaching a DL's target audience is a very specific type of usability that was employed for evaluation of The Glasgow Story (TGS) digitization project (Anderson 2007). The goal of TGS was one of social inclusion and lifelong learning—that is, to reach users who were not using “old” cultural heritage resources (Anderson, 366). Lifelong learners were defined as “any adult learner, irrespective of their life stage, who accessed content on the history of Glasgow for personal development, interest, knowledge, structured or unstructured learning,” and content for the digital collection was selected to meet this demographic (Anderson, 368). A comprehensive evaluation of the success of TGS in meeting these users' needs was developed including both qualitative and quantitative data collection in two stages each of formative and summative evaluation (Anderson, 371). The evaluation process revealed as much about the processes' shortcomings as it did about the research questions. Benchmark data for lifelong learners was not established prior to the evaluation, making it difficult if not impossible to determine if TGS is reaching its target demographic (Anderson, 376). The relative age of TGS's users was slightly older than the population of Glasgow as a whole, which could imply that the target population was being reached, but it is unknown whether users who completed the evaluation feedback form were from Glasgow and the small, self-selected sample may not be representative of all TGS users (Anderson, 376). Ethnic minorities, a group assumed to be socially excluded, were represented by a slightly smaller percentage among users than are represented in the total Glasgow population, so this group does not appear to be reached successfully by TGS (Anderson, 376). Additional flaws in

data collection and methodology limited the usefulness of the data collected (Anderson, 381). The final findings of TGS's evaluation concluded that a "modular set of metrics and evaluation instruments" is needed for institutions to adapt for evaluation so that they are not reinventing the wheel every time a digitization project is assessed (Anderson, 384).

The difficulties in information retrieval of different types of materials in DLs, in this case digitized newspapers, is another more specified attempt at addressing usability (Reakes and Ochoa 2009). The University of Florida's Florida Digital Newspaper Library (FDNL) and the National Digital Newspaper Project (NDNP)/Library of Congress Chronicling America digital newspaper collection were subjected to usability testing in 2008 as part of a NDNP grant requirement (Reakes and Ochoa, 96). Challenges specific to newspaper digitization include the complexity and variety of layouts from paper to paper, inconsistencies in section titles, large image sizes, difficulties with optical character recognition processes, metadata creation, and page segmentation, and are often the result of DAM limitations (Reakes and Ochoa, 94). Both the FDNL and Chronicling America interfaces were assessed by users who completed a number of scenarios and filled out pre- and post-test questionnaires. The results of the usability test were that the resource homepages left the most room for improvement (Reakes and Ochoa, 100). Some of the issues known to newspaper digitization mentioned previously were expressed by users even though some, such as article level retrieval, were intentionally disregarded in testing (Reakes and Ochoa, 104). The researchers concluded that a broader cross-section of newspaper digitization collections needed to be assessed to determine more concrete results, and that usability testing can actually lead to savings for institutions by

pointing software programming development towards user-requested functionality (Reakes and Ochoa, 108).

General evaluations of the user experience with institutions' specific DLs comprise a small section of the literature. Researchers at Colorado State University conducted usability testing on the Colorado State University Libraries' Digital Collections Web site and the Western Waters Digital Library to determine ease of use through real-life searching and users' perceptions of ease of use (Zimmerman and Paschal 2009, 229). While the users had some problems completing tasks using both of the DLs, they rated their experiences higher than the researchers expected (Zimmerman and Paschal, 236). A similar evaluation of the New Jersey Digital Highway used a Web-based online survey to assess the usefulness of the site from the perspectives of general users, educators, and cultural heritage professionals (Jeng 2008, 18). The results of the survey were generally positive, leading to new tool building to help more institutions contribute to the collection (Jeng 22-23).

Finally, a few institutions have published user studies that specifically look at the demographic population of DL users. A 2006 publication from the staff of the California Digital Library (CDL) looked at four years of user input and usage logs (Lack). The article includes the top ten themes found from user input at CDL as well as a user-centered product development model. Staff at East Carolina University investigated typical users of special collections with the goal of creating an interface that meets the search needs of both undergraduates and humanities researchers. The study found that humanities researchers want entire collections digitized, while undergraduates are accustomed to item-level description (such as that seen in library catalogs and databases) and want direct links to digitized materials, including in the

finding aid (Gueguen 2010, 98-99). East Carolina University's J.Y. Joyner Library designed a new DAMS to meet the needs of both user groups, which includes broad thematic collections, collection templates, subject clouds, hypertext links in item records to subjects and collection names, search facets, and user-generated content areas (comments and tagging). They also redesigned their EAD (Encoded Archival Description) stylesheet and navigation so that each EAD now includes a tab for all digitized objects from the collection (instead of individual links per item) (Gueguen).

### **Web Analytics**

The second most prevalent topic in DL assessment literature concerns the use of Web data to analyze usage and search patterns. The advantages of analyzing Web data include being able to increase knowledge about which links or items are being viewed the most in a DL, analyze the usage of finding aids/EADs in a DL to help prioritize which collections to digitize, evaluate and optimize online outreach attempts, measure the effectiveness of descriptive metadata, determine user demographics, and--when combined with other tools and methods--give a holistic view of users. There are a number of methods and tools for collecting Web data, including combining "user panels and browser logging tools to track sample WWW user populations; collecting network traffic data directly from ISP servers; and using site-specific server log parsers or page tagging technologies to measure traffic through a particular site" (Khoo et al. 2008, 375). Simple page view counts are unreliable because search engine spiders and robots may or may not be excluded in Web stats; caches may prevent visits from being logged in the server file; and dynamically generated pages made up of multiple elements may get logged as multiple page views (Voorbij 2010, 268). Page tagging is by far the most prevalent

technique found in the literature through the use of tools like AWStats, Webalizer, Urchin, Google Analytics, and Omniture (Khoo et al., 375-376). Page tagging eliminates the issues discussed above regarding page view counts, but page-tagging does not work for non-HTML pages or for users who do not have JavaScript enabled and has issues tracking users using IP addresses (Voorbij, 269).

The most common metrics reported are “the number of visitors to a site; the time and date of their visit; the geographical location of their IP address; whether they arrived via a search engine, bookmark, or link; the page(s) they enter and leave the site; the page(s) they viewed; time spent on individual pages; operating system; and monitor and browser configurations” (Khoo et al. 2008, 376). Because each of the tools previously mentioned measures these metrics in different ways, it is essential that libraries conducting Web usability report which tool was used (Khoo et al., 376). Additionally, the development of a sustainable Web metrics program necessitates that adequate resources be provided and maintained, including safe and stable access to Web servers, sufficient and capable staff, triangulation of the Web analytics with other data sources (like usability findings and the other methods detailed in this paper), and the knowledge that Web metrics results “are NOT ambiguous” (Khoo et al., 377-383). These ideas are echoed in a 2010 study on the use of Web statistics by libraries, archives, and museums in the Netherlands, where Google Analytics was the most commonly used tool, and visits and visitors were the most commonly reported data (Voorbij).

The remainder of the literature is comprised of case studies involving the use of page-tagging technologies to analyze DL usage, the most common of which is the free service Google Analytics. The report from a 2004 two-day workshop to develop a Web analytics strategy for

the National Science Digital Library (NSDL) resulted in the use of the page-tagging method to answer the questions “Who is coming to NSDL and to its individual sites? What do these users want from the sites? What works and does not work for these users?” (Sumner et al. 2004). At the beginning of the NSDL project, individual projects maintained their own metrics, resulting in a lack of standardization in what data was being collected and how it was being reported (Khoo 2006, 1). The “Core Integration” program, begun in 2005, required that the projects use Omniture for their Web analytics (though individual projects were still free to use their own tool of choice as well) (Khoo, 1). The NSDL Metrics Working Group now recommends that projects use Google Analytics for automatic reporting to the NSDL (Lightle et al. 2010, 3).

Other reports in the literature involve the usage of Google Analytics and DLs at the Illinois Harvest Portal (K. Hess 2012), EADS at East Carolina University’s Special Collections Division (Custer 2013), the Ball State University Digital Media Repository (Szajewski 2013), Washington State University Libraries’ EADs in the Northwest Digital Archives (NWDA) (O’English 2011), EADs at the University of Illinois (Prom 2011), the Gjenvick-Gjønvik Archives DL (Ament-Gjevick 2012), and multiple platforms (the library Web site, LibGuides, and CONTENTdm) at DePaul University (M. Hess 2012). Finally, a 2014 study utilized Google Analytics for both the NSDL and Opening History to analyze search behaviors of large populations rather than to assess the use of a single repository (Zavalina and Vassilieva 2014). Among other findings, the study determined that there are definite differences in the ways that users search STEM (science, technology, engineering, and mathematics) DLs versus cultural heritage DLs. The study recommended that STEM DLs should have more faceted search options and limits and should indicate object and concept in metadata, while cultural heritage DLs

targeted at educators, students, and researchers of history and the social sciences need to include more item attributes, including persons and places, in metadata (Zavalina and Vassilieva, 95-96).

As a whole, Web analytics can be seen as a method to develop enhancements to the architecture, metadata, and content of a DL to improve both user experience and success. The remainder of the literature on DL assessment is varied, but a few core subjects stand out.

### **Altmetrics**

Although primarily utilized to evaluate the impact of scholarly publications online, altmetrics have been mentioned as a potential tool for evaluating the usage of DLs as well. As scholars' general workflow moves increasingly to the Web, alternatives to traditional means of evaluating the quality of published resources (peer-review, citation counting, and JIF (journals' average citations per article)) are necessary to reflect changes in academic publishing and scholars' access to and use of information (Priem et al. 2010). Altmetrics can, in effect, "crowd-source peer-review," as the impact of a resource can immediately be evaluated through bookmarks, citations, mentions, and other methods of sharing information online (Priem et al.). The trick to utilizing altmetrics is in acknowledging that "buzz" is not necessarily equivalent to impact, but that used in combination with other methods of analysis they can be beneficial in measuring the reach of a resource (Priem et al.). Altmetrics are potentially useful for DLs because they focus not just on citations but on how data (or objects) are being reused.

One area in which altmetrics measure use differently from traditional scholarly metrics is in the realm of social media. DLs can use social media not just to promote their digital collections, but to interact with users in a meaningful way and therefore learn about the use

and usability of DLs. RSS search feeds, Twitter Search, Delicious, and Technorati™ are all tools that not only help DL managers find where their content is being mentioned or reused, but also to see how the online public is talking about a specific subject (Schier 2011). Allowing users to post both their praise and criticism of DLs establishes transparency and trust with users while also creating an open conversation—and valuable metrics—about the DL and its content (Schier). Some other tools that may be useful for DLs include:

- Scholarly reference: Bookmarking, shares and recommendations from CiteULike, Zotero, Mendeley
- Mass media mentions: NYTimes, BBC, The Washington Post
- Social media mentions: Twitter, Facebook, Delicious
- Data and code usage: Dryad, GitHub
- Component mentions: SlideShare, Figshare

(Groth and Taylor 2013)

The tools Altmetric and ImpactStory allow usage tracking via digital object identifiers (DOIs) or other IDs, including data for individual Web pages; this has potential for building a report of usage of items from a DL (Groth and Taylor). Though relatively new to the academic canon, especially as it related to DLs, altmetrics are conceivable tools for evaluating the impact of individual resources in DLs.

## **Reuse**

Similar to altmetrics, reuse studies determine whether DL materials are being reused online. Two studies have focused on the reuse of images using Reverse Image Lookup technologies (RIL). Standard natural metrics (like YouTube views and Web site hits) are not

necessarily applicable to digital images if they are used outside of context, nor are text queries particularly accurate in image retrieval (Kousha, Thelwall, and Rezaie 2010, 1735). Researchers used TinEye, a RIL search engine, to analyze the reuse of unique, free, open access images from NASA (Kousha, Thelwall, and Rezaie, 1735). The search engine was analyzed to see if it could identify online copies of academic images, determine common motivations for copying academic images online and see how that information could be used in regards to research impact, and whether or not different types of images affected TinEye's retrieval abilities (Kousha, Thelwall, and Rezaie, 1736). TinEye was found to be somewhat accurate as it could find exact matches for images even if they had been cropped, resized, or edited, but it sometimes retrieved several similar images from a single Web site, or reported repetitious results if a Web site contained several sizes of the same image (Kousha, Thelwall, and Rezaie, 1737). After deleting duplicate results, the reuse of the images was classified according to where and how the image appeared. The results of the classification showed differences in trends found in academic publications reuse. Time had little effect on reuse, unlike with academic publications; older images were not used more, possibly because Web search favors newer content (Kousha, Thelwall, and Rezaie, 1738-1739). Few images were found in research publications, but that could be because TinEye doesn't index PDF, doc, or PS files, and many research publications are not available on the open Web (Kousha, Thelwall, and Rezaie, 1738-1739). Over one third of the images were used for informal scholarly or educational communication, one fourth for backgrounds and layouts, and just less than one fourth for navigational illustration (Kousha, Thelwall, and Rezaie, 1738-1739). Secondary studies of visual arts images and biology and medical images were also undertaken. While TinEye allowed for

easy and free reverse image lookup, there were some limitations. TinEye's indexing policy is unknown, so the currency of results as well as what sites are indexed is in question (Kousha, Thelwall, and Rezaie, 1741). Limits on what file types can be searched (HTML only) also affected results, as did the quality of the images uploaded (the best matches found were with high quality images) (Kousha, Thelwall, and Rezaie, 1741).

A similar study contrasted RIL of images from The National Gallery (UK) searched in both TinEye and Google Image Search, using Content Analysis to discover contexts for reuse, and then triangulated with Google Analytics and stats from a commercial ISP firm (Kirton and Terras 2013). As in the previous study, the researchers found TinEye's indexing to be a limitation; many results did not seem to actually contain the image, suggesting that TinEye's crawl was outdated (Kirton and Terras). Google Images crawls more Web sites and therefore had a greater results set, but it is less transparent as it self-regulates and removes similar results (Kirton and Terras). Triangulating the results of the RIL with Web statistics from Google Analytics and HitWise showed that the most accessed images on the National Gallery Web site were also the most reused elsewhere online (Kirton and Terras). Also, the reuse of images elsewhere online directs traffic back to the National Gallery Web site, showing potential venues for outreach (Kirton and Terras). RIL is time-consuming using these free technologies, so it is best used for providing information on targeted parts of a collection (Kirton and Terras). Still, the researchers determined that the freer the license, and the more reuse of digital content, the more the original institution will benefit (Kirton and Terras).

Another method for determining the context for image reuse was discussed in regards to the University of Houston Digital Cart Service (the home-grown digital image delivery service

for DL). The Digital Cart Service (DCS) was developed in collaboration with their IT department to let users request 600dpi images from the university's CONTENTdm collections for free delivery through email (Reilly and Thompson 2014, 197). DCS records patron-provided data including name, date, image file name, affiliation, and description of use, creating "ultimate use" data, or the purpose for which users are requesting high resolution images (Reilly and Thompson, 198). The researchers found that use purpose varied by user group (Reilly and Thompson, 204). Users were accessing images for reuse in publications (both popular culture products and scholarly), research (personal, scholarly, industrial), and artwork (Reilly and Thompson, 207-208). Knowledge about ultimate use has implications for metadata creation, system design, marketing and promotion, and content selection (Reilly and Thompson, 209). The types of uses for images also led the researchers to believe that concepts are more helpful in image description than attributes (e.g., "color" or "24-bit"), and that the incorporation of user-generated content into metadata could be beneficial (Reilly and Thompson, 209).

The importance of descriptive metadata in image retrieval was reinforced in a use study involving journalists and historians, faculty and current and former students at Dalhousie University. The authors identified two primary approaches to image retrieval: manually created metadata and automated techniques; the authors found a combination of the two is the most successful method for delivering images successfully (McCay-Peet and Toms 2009, 2417). Users reported that they retrieved images for illustration purposes more often than for informational purposes (McCay-Peet and Toms, 2422-2423).

Finally, link analysis can also be conducted to determine where users are reposting links to a DL. While hyperlinks viewed out of context may not necessarily denote an endorsement,

they still help point to the general reach of a resource. The Toolkit for the Impact of Digitised Scholarly Resources (<http://microsites.oii.ox.ac.uk/tidsr/>) provides a number of qualitative and quantitative tools and methods for determining reach, including link analysis tools. A study of the usage and impact of five specific digitized scholarly resources (Histpop – Online Historical Population Reports; 19th Century British Library Newspapers (phase one); British Library Archival Sound Recordings (phase one); 18th Century Official Parliamentary Publications Portal 1688-1834 at the British Official Publications Collaborative Reader Information Service; and the Wellcome Medical Journals Backfiles) used Webometric Analyst, formerly LexiURL Searcher, for an analysis comparing the links to each of the digital resources to a set of comparator Web sites (Eccles, Thelwall, and Meyer 2012, 513-514). The advantages to Webometrics are that data is easy to acquire as is comparison with other sites, as a result, benchmarking is possible. The disadvantages are that hyperlink creation is not necessarily an endorsement, and links may be created or duplicated automatically as part of the Web design and therefore may not be true examples of intentional link reposting (Eccles, Thelwall, and Meyer, 513).

Examining the reuse of DL materials points to the reach and possible usefulness of digital collections but can be time-consuming to conduct and difficult to quantify as representative of a successful DL. Image and link reuse studies are still relatively scarce in the professional literature, but an increase in the quantity of these studies will allow other institutions to determine if this type of assessment is cost- and time-efficient for their DL assessment plan.

### **Cost Benefit**

Even scarcer in the literature are examples of studies to determine the overall costs of a DL, known as cost benefit analysis. Cost benefit analysis is integral to the evaluation of DLs because it provides financial justification for the digitization and sustainability of collections. But calculating the total cost of a DL and contrasting that with the money “saved” by creating the project is not always a cut-and-dried process as most DLs do not charge for use. Costs incurred in the creation of the DL include recurring (maintenance of the project) and nonrecurring (initial implementation) costs, and both hard (for example, purchasing software) and soft (such as the labor involved in implementing said software) costs (Cervone 2010, 77). An essential part of a cost benefit analysis is the “payback time” or “breakeven point” when the project is “paid for” (Cervone, 77). The simplest model involves only simplified costs and benefits and does not take into account intangible costs and benefits, like the prestige in self-hosting a DL versus hosting by a vendor despite possible higher labor costs for the host institution (Cervone, 78).

The digital library team at Portland State University’s Millar Library asked the question, “What kind of cost responsibilities does a library assume when building a digital library?” in relation to the Oregon Sustainable Community Digital Library (Hickox et al. 2006, 52). The cost benefit analysis allowed the institution to determine what percentage of costs were spent on what phases of the DL program: pre-processing and research and administration (45 percent each, or 90 percent total) and Outreach, Cataloging, Servers/Storage, and Design (10 percent) accounted for the pre-digitization phase of the project (Hickox et al., 59). The authors caution that costs vary widely due to personnel salaries and the types and conditions of the materials being digitized (Hickox et al., 61).

Another study involved cost benefit analysis at the Triangle Research Libraries Network to determine whether quality control visual checks were cost efficient in large-scale digitization projects. The results showed that 85 percent of the time was spent scanning and 15 percent on quality control with visual scans of all items (Chapman and Leonard 2013). Only 0.4 percent of scans had errors, and only 0.1 percent had critical errors; production could have increased by 18 percent if the quality control checks had not been performed, and this would have had little effect on the overall quality of the project (Chapman and Leonard). The authors also looked at which materials caused the most critical errors and determined that quality control checks could be limited only to these types of items, or only during the training of new scanning technicians (Chapman and Leonard).

Cost benefit analysis is a very useful method for determining the relative worth of a DL project. As in-kind contributions are relatively unique to different institutions, there will likely be some level of customization for each new DL's cost benefit analysis. Still, continued contributions to the professional literature in this area could lead to the development of general rubrics and tools for estimating the hidden costs of DL projects, which in turn can help institutions weigh the possible benefits of DLs against the total expenditures.

### **Holistic Approach to DL Assessment**

Finally, the most comprehensive way to evaluate DLs should involve multiple methods, or a holistic approach. The different methods of analyzing DLs mentioned previously in this article should be combined to get a larger picture of a DL's successes and shortcomings. The scholarly publications in this area comprise nine percent of the total literature explored in this

paper but should be explored by DL stakeholders as some of the most useful and comprehensive examples of DL assessment.

A 2010 study surmised that traditional information retrieval methods tend to be used for evaluating DLs: “Few metrics reflect unique DL characteristics, such as variety of digital format. And few address the effects of a DL at higher levels, including the extent to which a DL fits into or improves people’s daily work/life” (Zhang 2010, 88). Existing models of evaluation utilize criteria for content, technology, interface, service, user, and context (Zhang, 88). Of these six levels, the body of research towards evaluation of digital content is especially weak (Zhang, 88-89). When digital content is evaluated, it is broken up into four categories: digital objects, metadata, information, and collection (Zhang, 89). Digital objects are the most specific to DLs and are assessed with DL-specific criteria like fidelity and suitability to the original artifact (Zhang, 89). As noted earlier in this review of the literature, there are an abundance of usability studies, and the interface is the most frequently evaluated element and has the most defined criteria (Zhang, 89). The literature also contains frameworks and models for benchmarking evaluations (for comparison against other DLs to measure success), but these are not specifically related to DLs (Zhang, 89). Zhang therefore sought to determine what criteria are specific to evaluating DLs, which of these criteria are the most important, and how they can be presented in the most meaningful ways (Zhang, 90). The Rutgers library Web site, which includes access to digital collections, was used as a test subject. The proposed model included context, content, technology, interface, users, and service; each includes core criteria as well as group-based criteria from five user groups (general, researcher, librarian, developer, and administrator) (Zhang, 99). Users were found to be most interested in accessibility to content and

sustainability of the DL, with interaction with the content and DL performance prioritized next (Zhang, 107). The subsequently developed model was well-received in the verification stage, but Zhang concluded that it needed to be further tested in a more diverse setting before it could be truly adopted (Zhang, 104).

The University of Alabama used a dual cost benefit/usability model to evaluate the feasibility of using the Archivists' Toolkit plugin to add batches of digital content directly to EADs. The analysis compared costs between the normal and newly proposed workflows by calculating the time averages for work (DeRidder, Presnell, and Walker 2012, 151). The new method saved 390.93 minutes/per 100 scans per process step; overall, the new method saved \$78,000 over the original workflow (DeRidder, Presnell, and Walker, 157). A usability study was then conducted with 20 participants doing four known-item searches in two collections (one collection created using the old workflow, and one with the new). The old collection was easier and quicker to use, except for users without previous digital collection experience; the researchers were therefore able to surmise that EADs with digital content are more suitable for scholars than students (DeRidder, Presnell, and Walker, 169). An evaluation like this allows libraries to weigh the benefit of saving money directly with the effect a new process might have on the user experience.

DL evaluations can also be combined with traditional library service assessment criteria for holistic evaluation. The University of South Florida Tampa Library developed a holistic assessment including data from Aeon (their material request and workflow management software), Desk Tracker, reading room patron surveys, Web site and digital collections usability testing, Web analytics, and Fedora Commons analytics (Griffin, Lewis, and Greenberg 2013). Utilizing multiple assessment methods allowed for improved Web navigation, optimized DL performance for Internet Explorer and mobile devices (determined to be heavily adopted by users), the

development of digitization priorities, and outreach opportunities based on little-used physical collections with high Web views (Griffin, Lewis, and Greenberg, 234-235).

Finally, the previously mentioned Toolkit for Digital Scholarly Resources (TIDSR) was developed in 2009 by the Oxford Internet Institute through funding from JISC to aid institutions in using open tools to quantitatively assess the “footprint” of a digitized resource and qualitatively answer questions about its value (Hughes 2014). The TIDSR Web site has links to many case studies from institutions that have used the various tools recommended. One example sought to analyze the following information for the Welsh Journals Online collection of digitized scholarly resources: the number of new and returning users; who the user communities are; how users locate and access the Web site; and whether or not there is evidence of use of the collection in scholarship (Hughes). By analyzing referral data and who the user communities were, the study determined that the primary users were genealogists and that the collection needed to be promoted better to academics (Hughes). The citation analysis resulted in few citations for digitized journals but some for print; after contacting known users, it was found that some had used a print citation erroneously and would appreciate a citation tool (Hughes).

Holistic evaluations of DLs are the most complete method of DL assessment and should be viewed by DL stakeholders as a best practice. The possible combinations of DL assessment methods and tools are vast, so the literature in this area is important for institutions to determine which types of evaluations can be combined to achieve the greatest overall picture of a DL’s success.

## **Conclusion**

Professional literature about the assessment of digital libraries reflects a growing interest in both improving the user experience and in justifying the creation of digital collections to multiple stakeholders. While previous reviews have also found that usability or user studies and Web analytics are the most prevalent subjects in the DL literature (Liew 2009, Todd-Diaz and O'Hare 2014), newer areas of study including altmetrics, which is growing in popularity among analyses of scholarly publications online, and cost-benefit analysis, essential for the justification of DLs especially in times of budget limitations, seem ripe for growth in the professional research. Reuse analysis is also a more recent area of study and needs to be further explored to identify whether it is a useful method of evaluation for a wide variety of institutions. Holistic evaluations too are necessary additions to the canon especially if the literature can address the needs of varying types and sizes of institutions.

Other areas which are mentioned in the evaluation of other digital resources include benchmarking and sustainability; while not explored here, these are areas for growth in DL assessment scholarship. While some scholarship mentions the need for benchmarks, or measurements used to select a "good" reference for comparison, there is little written on how individual institutions can set benchmarks for their own collections. Similarly, some surveys of DL sustainability practices among institutions have been published, but there is a lack of scholarly output in the area of evaluating a DL's potential for sustainability. As they are important contributions to the DL literature, benchmarking and sustainability may see an increase in publications in the DL literature in the coming years.

Reviewing the scholarship regarding assessment of DLs is an integral step for DL stakeholders in determining how and why DL evaluation is necessary. DL stakeholders can use

existing studies to inform their own evaluation practices and, hopefully, then also contribute to the DL literature. The greater the canon of rigorous and honest evaluations of DL available, the greater the possibility that DL projects will be created thoughtfully and strategically and then maintained and modified to best meet user needs.

## References

- Ament-Gjevick, Paul. 2012. "Using Web Analytics—Archival Websites". Clayton State University.  
<http://www.archivalstudies.com/Documents/ARST5110-AmentGjenvick-UsingWebAnalyticsResearchPaper.pdf>.
- Anderson, Ian G. 2007. "Pure Dead Brilliant?: Evaluating The Glasgow Story Digitisation Project." *Program* 41 (4): 365–85.  
doi:<http://dx.doi.org.ezproxy.loyno.edu/10.1108/00330330710831585>.
- Cervone, H. Frank. 2010. "Using Cost Benefit Analysis to Justify Digital Library Projects." *OCLC Systems & Services* 26 (2): 76–79. doi:10.1108/10650751011048443.
- Cervone, Frank. 2014. "Evidence-Based Practice and Web Usability Assessment." *OCLC Systems & Services* 30 (1): 11–14. doi:10.1108/OCLC-10-2013-0036.
- Chapman, Joyce, and Samantha Leonard. 2013. "Cost and Benefit of Quality Control Visual Checks in Large-Scale Digitization of Archival Manuscripts." *Library Hi Tech* 31 (3): 405–18. doi:<http://dx.doi.org.ezproxy.loyno.edu/10.1108/LHT-01-2013-0002>.
- Comeaux, David J. 2008. "Usability Studies and User-Centered Design in Digital Libraries." *Journal of Web Librarianship* 2 (2/3): 457–75.
- Custer, Mark. 2013. "Mass Representation Defined: A Study of Unique Page Views at East Carolina University." *American Archivist* 76 (2): 481–501.
- DeRidder, Jody, Amanda Presnell, and Kevin Walker. 2012. "Leveraging Encoded Archival Description for Access to Digital Content: A Cost and Usability Analysis." *American Archivist* 75 (1): 143–70.

- Dickson, Maggie. 2008. "CONTENTdm Digital Collection Management Software and End-User Efficacy." *Journal of Web Librarianship* 2 (2/3): 339–79.  
doi:10.1080/19322900802190852.
- Eccles, Kathryn E., Mike Thelwall, and Eric T. Meyer. 2012. "Measuring the Web Impact of Digitised Scholarly Resources." *Journal of Documentation* 68 (4): 512–26.  
doi:http://dx.doi.org.ezproxy.loyno.edu/10.1108/00220411211239084.
- Griffin, Melanie, Barbara Lewis, and Mark I. Greenberg. 2013. "Data-Driven Decision Making: An Holistic Approach to Assessment in Special Collections Repositories." *Evidence Based Library & Information Practice* 8 (2): 225–38.
- Groth, Paul, and Dario Taylor. 2013. "Helping Scholars Tell Their Stories Using Altmetrics." *Library Connect*. <http://libraryconnect.elsevier.com/articles/supporting-users-organizations/2013-08/helping-scholars-tell-their-stories-using-altmetrics>.
- Gueguen, Gretchen. 2010. "Digitized Special Collections and Multiple User Groups." *Journal of Archival Organization* 8 (2): 96–109. doi:10.1080/15332748.2010.513324.
- Hess, Kirk. 2012. "Discovering Digital Library User Behavior with Google Analytics." *The Code4Lib Journal*, no. 17 (June). <http://journal.code4lib.org/articles/6942>.
- Hess, M. Ryan. 2012. "Web Analytics: Using Evidence for Improvement - Über Analytics: Customizing Google Analytics to Track Multiple Library Platforms." In <http://works.bepress.com/mryanhess/9>.
- Hickox, Chessa Grasso, Rose M. Jackson, Gary W. Markham, and Christopher N. Cox. 2006. "Going Broke, Going Digital: A Preliminary Cost Analysis of Building a Digital Library." *Internet Reference Services Quarterly* 11 (1): 51–66.

- Hughes, Lorna M. 2014. "Live and Kicking: The Impact and Sustainability of Digital Collections in the Humanities." In *Studies in the Digital Humanities*. Sheffield: HRI Online Publications. <http://www.hrionline.ac.uk/openbook/chapter/dhc-hughes>.
- Jeng, Judy. 2005. "Usability Assessment of Academic Digital Libraries: Effectiveness, Efficiency, Satisfaction, and Learnability." *Libri: International Journal of Libraries & Information Services* 55 (2/3): 96–121.
- Jeng, Judy. 2008. "Evaluation of the New Jersey Digital Highway." *Information Technology and Libraries* 27 (4): 17–24.
- Khoo, Michael, Mimi Recker, Joe Pagano, Bart Palmer, Anne Washington, and Robert Donahue. 2008. "Using Web Metrics to Analyze Digital Libraries." In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*, 375–84. [http://www.ist.drexel.edu/faculty/mkhoo/docs/09\\_jcdl.pdf](http://www.ist.drexel.edu/faculty/mkhoo/docs/09_jcdl.pdf).
- Khoo, Mick. 2006. "NSDL Web Metrics." *National Science Digital Library Project Archive*. [http://nsdl.library.cornell.edu/websites/comm/uploads/Web\\_metrics\\_NVC\\_4-25-06.pdf](http://nsdl.library.cornell.edu/websites/comm/uploads/Web_metrics_NVC_4-25-06.pdf)
- Kirton, Isabella and Melissa Terras, "Where Do Images of Art Go Once They Go Online? A Reverse Image Lookup Study to Assess the Dissemination of Digitized Cultural Heritage." In *Museums and the Web*, eds. N. Proctor & R. Cherry (Silver Spring, MD: Museums and the Web, 2013). <http://mw2013.museumsandtheweb.com/paper/where-do-images-of-art-go-once-they-go-online-a-reverse-image-lookup-study-to-assess-the-dissemination-of-digitized-cultural-heritage/>

- Kousha, Kayvan, Mike Thelwall, and Somayeh Rezaie. 2010. "Can the Impact of Scholarly Images Be Assessed Online? An Exploratory Study Using Image Identification Technology." *Journal Of The American Society For Information Science & Technology* 61 (9): 1734-1744.
- Lack, Rosalie. 2006. "The Importance of User-Centered Design: Exploring Findings and Methods." *Journal of Archival Organization* 4 (1/2): 69–86. doi:10.1300/J201v04n01-05.
- Lightle, K., Almasy, E., Barbato, L., Clark, S., George, Y., Hsi, S., Lowe, C., MacKinney, P., McIlvain, E. (2010). *Best Practices for Metrics Collection and Analysis: Report of the NSDL Metrics Working Group*. 1-39. [http://nsdl.library.cornell.edu/materials/outreach\\_and\\_communications/BestPracticesMetrics-NSDL-Aug2010.pdf](http://nsdl.library.cornell.edu/materials/outreach_and_communications/BestPracticesMetrics-NSDL-Aug2010.pdf)
- Liew, Chern Li. 2009. "Digital Library Research 1997-2007: Organisational and People Issues." *Journal of Documentation* 65 (2): 245–66.  
doi:http://dx.doi.org.ezproxy.loyno.edu/10.1108/00220410910937606.
- Long, Holley, Kathryn Lage, and Christopher Cronin. 2005. "The Flight Plan of a Digital Initiatives Project, Part 2: Usability Testing in the Context of User-Centered Design." *OCLC Systems & Services* 21 (4): 324–45.
- McCay-Peet, Lori, and Elaine Toms. 2009. "Image Use within the Work Task Model: Images as Information and Illustration." *Journal of the American Society for Information Science and Technology* 60 (12): 2416–29. doi:10.1002/asi.21202.

- O'English, Mark. 2011. "Applying Web Analytics to Online Finding Aids: Page Views, Pathways, and Learning about Users." *Journal of Western Archives* 2 (1).
- Paterson, Lorraine, and Boon Low. 2010. "Usability Inspection of Digital Libraries." *Ariadne: A Web & Print Magazine of Internet Issues for Librarians & Information Specialists* 30 (63).
- Priem, Jason, Dario Taraborelli, Paul Groth, and Cameron Neylon. 2010. "Altmetrics: A Manifesto – Altmetrics.org." <http://altmetrics.org/manifesto/>.
- Prom, Christopher. 2011. "Using Web Analytics to Improve Online Access to Archival Resources." *American Archivist* 74 (Spring/Summer 2011): 158–84.
- Reakes, Patrick, and Marilyn Ochoa. 2009. "Non-Commercial Digital Newspaper Libraries: Considering Usability." *Internet Reference Services Quarterly* 14 (3/4): 92–113.
- Reilly, Michele, and Santi Thompson. 2014. "Understanding Ultimate Use Data and Its Implication for Digital Library Management: A Case Study." *Journal of Web Librarianship* 8 (2): 196–213. doi:10.1080/19322909.2014.901211.
- Schier, Robert A. 2011. "Digital Librarianship @ Social Media: The Digital Library as Conversation Facilitator." *D-Lib Magazine* 17 (7/8). doi:10.1045/july2011-schrier.
- Smith, Carol. 2006. "Multiple Cultures, Multiple Intelligences: Applying Cognitive Theory to Usability of Digital Libraries." *Libri: International Journal of Libraries & Information Services* 56 (4): 227–38.
- Sumner, Tamara, Casey Jones, Sarah Giersch, Michael Wright, Anita Coleman, and Laura Bartolo. 2004. "Developing a Web Analytics Strategy for the National Science Digital Library." *D-Lib Magazine* 10 (10). <http://www.dlib.org/dlib/october04/coleman/10coleman.html>.

- Szajewski, Michael. 2013. "Using Google Analytics Data to Expand Discovery and Use of Digital Archival Content." *Practical Technology for Archives* 1 (1).  
[http://practicaltechnologyforarchives.org/issue1\\_szajewski/#ftn10](http://practicaltechnologyforarchives.org/issue1_szajewski/#ftn10).
- Todd-Diaz, Ashley, and Sheila O'Hare. "If You Build It, Will They Come?" *Annual Review of Cultural Heritage Informatics: 2012-2013* (2014): 257-275.
- Voorbij, Henk. 2010. "The Use of Web Statistics in Cultural Heritage Institutions." *Performance Measurement & Metrics* 11 (3): 266–79.
- Xie, Hong (Iris). 2006. "Evaluation of Digital Libraries: Criteria and Problems from Users' Perspectives." *Library & Information Science Research* 28 (3): 433–52.  
doi:10.1016/j.lisr.2006.06.002.
- Xie, Hong (Iris). 2007. "Help Features in Digital Libraries: Types, Formats, Presentation Styles, and Problems." *Online Information Review* 31 (6): 861–80.  
doi:<http://dx.doi.org.ezproxy.loyno.edu/10.1108/14684520710841810>.
- Xie, Hong (Iris). 2008. "Users' Evaluation of Digital Libraries (DLs): Their Uses, Their Criteria, and Their Assessment." *Information Processing & Management* 44 (3): 1346–73.  
doi:10.1016/j.ipm.2007.10.003.
- Zavalina, Oksana, and Elena V. Vassilieva. 2014. "Understanding the Information Needs of Large-Scale Digital Library Users: Comparative Analysis of User searching. (Report)." *Library Resources & Technical Services* 58 (2).
- Zhang, Ying. 2010. "Developing a Holistic Model for Digital Library Evaluation." *Journal of the American Society for Information Science & Technology* 61 (1): 88–110.  
doi:10.1002/asi.21220.

Zimmerman, Don, and Dawn Bastian Paschal. 2009. "An Exploratory Usability Evaluation of Colorado State University Libraries' Digital Collections and the Western Waters Digital Library Web Sites." *Journal of Academic Librarianship* 35 (3): 227–40.