Measuring the Impact of Preserving Digital Assets

Kathleen Murray and Dreanna Belden; University of North Texas; Denton, Texas, USA

Abstract

The Portal to Texas History℠ is a gateway to humanities collections within the digital library of the University of North Texas (UNT) Libraries (http://texashistory.unt.edu). Currently, materials from more than 190 content partners are available and the number of partners continues to grow. While ever-increasing numbers of partners and assets are signs that digitally preserving and making resources Web-accessible is a desirable thing, universities, cultural heritage institutions, and funding agencies increasingly expect measurements that report the impact and value resulting from digitizing and preserving assets. Because the Portal is fairly unique in both the number and scope of its content partners, it serves as a good case study for measuring the impact of digitization for two key digital library stakeholder groups: content providers and users. This paper reports the initial findings of a study of the impact of digitizing assets, specifically: (a) a framework of impact areas and indicators, and (b) findings for the Portal's content partners and users.

Background

The Portal to Texas History

The Portal to Texas History℠ was launched in 2003 with collections from five content partners. Currently, materials from more than 190 partners are available and the number of partners continues to grow. The Portal archives and provides access to more than 150,000 digital objects, comprising over two million image files. The range of primary source materials includes maps, books, manuscripts, newspapers, diaries, photographs, and letters from the unique collections of Texas libraries, museums, archives, historical societies, genealogical societies, and private families. Materials primarily concern the 254 Texas counties, but there are items related to most of the states in the USA and to over 40 other countries. Because the Portal is fairly unique in both its number and scope of partners, it serves as a good case study for an impact study.

Digital Libraries & Value

Assessing the value of digitized assets continues to be a challenge. During the early period for digital libraries, efforts centered on infrastructure building and best practices and few evaluation studies were conducted. Only in the past five years have assessments begun to focus on usability and impact, while development of standards, methodology, and benchmarks for digital library assessment remain under development. [1]

Many studies focus primarily on usage as the main criteria to determine impact. Research reveals that digitization yields several results: digitized special collections materials will be used at a higher rate and increasing access can change the anticipated audience for these types of materials. [2] Digitization drives usage of libraries and is changing priorities as students and faculty prefer using digital resources [3], but the question of how digitized resources are meeting people's needs remains underexplored.

The area of digital libraries is evolving, and new research topics will emerge. [4] While the digital library community knows well the impacts outlined above, what has not been investigated in any depth is the actual meaning and value to individuals that underlies the impacts. What value do digitized assets represent to end users, both academic and lay researchers, and the institutions involved in digitization? How is the broad availability of these digital assets changing lives, influencing research, and finally transforming institutions engaged in these practices?

A JISC-funded project, *Inspiring Research, Inspiring Scholarship*, measured the impact and value of digitized resources in the United Kingdom in terms of learning, teaching, research, and society. Several significant areas of impact were identified and testimonies and case examples augmented the more quantifiable impacts. However, the overall conclusion was that better evidence of impact is needed and that the core question to answer is: "How has the digital resource delivered a positive change in a defined group of people's lives?" [5]

Writing from the perspective of the impact of a corporate library, Edgar asserts that a "problem often not addressed in the research done so far on the value and influence created by the corporate library is that the ultimate value provided by the corporate library has not been conceptualized holistically as customer value". [6] He goes on to say that, in addition to demonstrating its value to its own corporate clients and corporate operations, a library can expand the measurement of library value to include measures of value in terms of change for the better in external customers' lives.

Oakleaf asserts that "this reconceptualization of library value merits further investigation" for its applicability to academic libraries. [7] This suggests that it is reasonable for an academic library to measure its impact, including the value of the digitized resources in its collections, in terms of the positive changes made in the lives of external stakeholders.

Methods

This study was concerned with assessing the impact of digitizing resources on external stakeholders of The Portal to Texas History, specifically, content partners and users. In order to embrace the range of possible impacts, this study employed a mixed methods approach. Such an approach overcomes any limitations or bias resulting from any single measure of impact and allows the dimensions of impact to be refined from multiple perspectives and distinguished for different stakeholder groups.

A common framework of impact areas for digitized resources, as well as indicators for each area, was adapted from Tanner [5] (Table 1). The impact areas included: economic, social, educational, cultural, political, environmental, organizational, and operational. Using a common framework made it possible to conceptually organize findings from different data collection

activities and data sources, specifically, key informant interviews and user-submitted comments to the Portal system.

Table 1. Framework of Impact Areas and Indicators

Impact Areas	Indicators
Cultural	• Preservation and dissemination of cultural heritage • Existence value (a good thing to do; people like that this is being done) • Preservation of cultural heritage for future access
Economic	• Usage (unique use instances) • Revenue • Decreased costs of holding materials physically (floor space repurposing) • Funding increases • Risk management (avoidance of loss) • Return on investment (usage per item) • Tourism (promotion through social networks)
Educational	• Attendance at educational events (physical or virtual locations) • Feedback from educators • Richer and more accessible lesson plans and curricula • Outputs from higher education digital humanities research and digital scholarship
Environmental	• Avoidance of travel (economies of gasoline usage) • Elimination of physical curation
Operational	• Increased effectiveness in materials management • Efficient fulfillment of duplication requests • Staffing impact • No-cost outsourcing of resource preservation
Political	• Prestige value (enhanced public profile; improved visibility) • Marketing strategy (branded online product versus a physical item)
Social	• Sharing via social networks (facebook, twitter, blogs) • Increased access (search engines, Google) • Activities pursuant to organizational mission

In the initial stage of the project, researchers investigated the impact of digitizing resources on content providers by conducting key informant interviews and analyzing Portal usage data. The impact on users was investigated through citation analysis and content analysis of a Portal log of user-submitted comments. Both quantitative and qualitative methods were used to analyze the data.

Key Informant Interviews

A questionnaire to assess and measure the indicators for each impact area in Table 1 was created and reviewed for clarity and face validity by informational professionals working in digital libraries. The questionnaire was revised and finalized based on the feedback received.

Key informant interviews with five Portal partners were conducted in person, and the interviews were recorded, transcribed, and content analyzed to identify impact areas, as well as other common themes or categories that emerged. Interview participants came from a variety of institutions representative of the Portal's content partners: a museum, public library, archives, public academic library and private academic library.

Citation Analysis

The goal of the citation analysis was to identify Portal assets cited in scholarly works. This was done via structured searches of the indexed content in Google Books, Google Scholar, JSTOR, and Academic Search Premier. The resulting citation list was content analyzed to identify and quantify three dimensions: publication discipline, publication type, and the number of citations per year.

Content Analysis of User Feedback

Content analysis was conducted on the Portal's historical log of user-submitted comments. The Portal site provides a feedback form on every page and users submitted 3,276 comments between October 13, 2005 and November 30, 2011. Two researchers analyzed the content using the impact areas and indicators framework (Table 1). Additional categories and indicators were also identified and used in the analysis: Reference questions, Enhancement suggestions, Corrections (including metadata and other data additions), and Technical issues.

Researchers assigned 3,511 categories to the comments. Of these 2,422 were in agreement between the researchers and 1,089 differed. The agreements were further analyzed to determine the number and percentage of comments in each impact area and other category.

Usage Data

UNT Libraries captures usage data for all objects within the Portal. Researchers analyzed the aggregate usage data for each of two years for which 12-months of data were available.

UNT Libraries defines a single "use" as the aggregation of all interactions that a single IP address has with a digital object (e.g., a photograph or map) within a 30-minute period. Uses are aggregated at various levels (i.e., for content partners, collections, and the Portal system) to represent usage at these levels for a given period of time. The resolution used for all statistics is one day, with the number of uses for each object recorded per day.

Discussion of Results

Impact on Portal Content Partners

"The top three impacts… [a] accessibility, ease of reproduction; . . . [b] little Bobby and little Susie can download an image for their use in a paper for educational purposes; and [c] the other kind of intangible thing is transparency in government

because we are not sitting on our materials - they are accessible to everyone."
   -- Archivist

Key Informant Interviews

"So if we ask first about putting a certain group of materials up ... that will be used and useful . . . then what we need to find out after that is: [Are they] being used? And we haven't asked."
-- Academic Library Director, Private Institution

The interviews with key informants were analyzed using the impact areas in the common framework. Figure 1 illustrates the number of statements ($N = 220$) for each of the impact areas by partner type.

Cultural Impact: "The ease of access has allowed us to get things in local news, local media outlets, national news; when we made national news on the Bonnie and Clyde materials, that was

something a producer in New York was able to look at virtually, without having to send somebody over to look at the originals."
   -- Archivist

Of the total value statements, 36% fell into the cultural impact area, which encompasses the value digitization brings to preservation and dissemination of cultural heritage; existence value – this is a good thing to do; and preservation for future access. Access emerged as perhaps the most important aspect for

partners and was a major driver in their decision to contribute materials. Access and preservation form a key part of each of these institutions' missions, and content partners view digitization as a means of supporting this core mission.

Economic Impact: "Certainly by having this material digitized, we have, and in a secure location, we've limited that fact that if a tornado hits this building that we're sitting in and rips the top off and the archives is flooded, and so forth, at least a few things are safely ensconced elsewhere."
– Academic Library Director, Private Institution

Partner statements about economic factors resulting from digitization – factors such as digital usage counts, increased revenue, risk management/loss avoidance – encompassed 21% of the all value statements. Most partners noted that having a digital surrogate located elsewhere was an important strategy for avoiding loss of physical materials. None of the partners interviewed identified any significant revenue impacts from digitization activities; however, some noted an increase in the number of images licensed or reprinted, which was attributed to increased discoverability of the items.

Political Impact: "The city ...the governing body of the city was really excited about the partnership and having the collections accessible."
– Museum Director

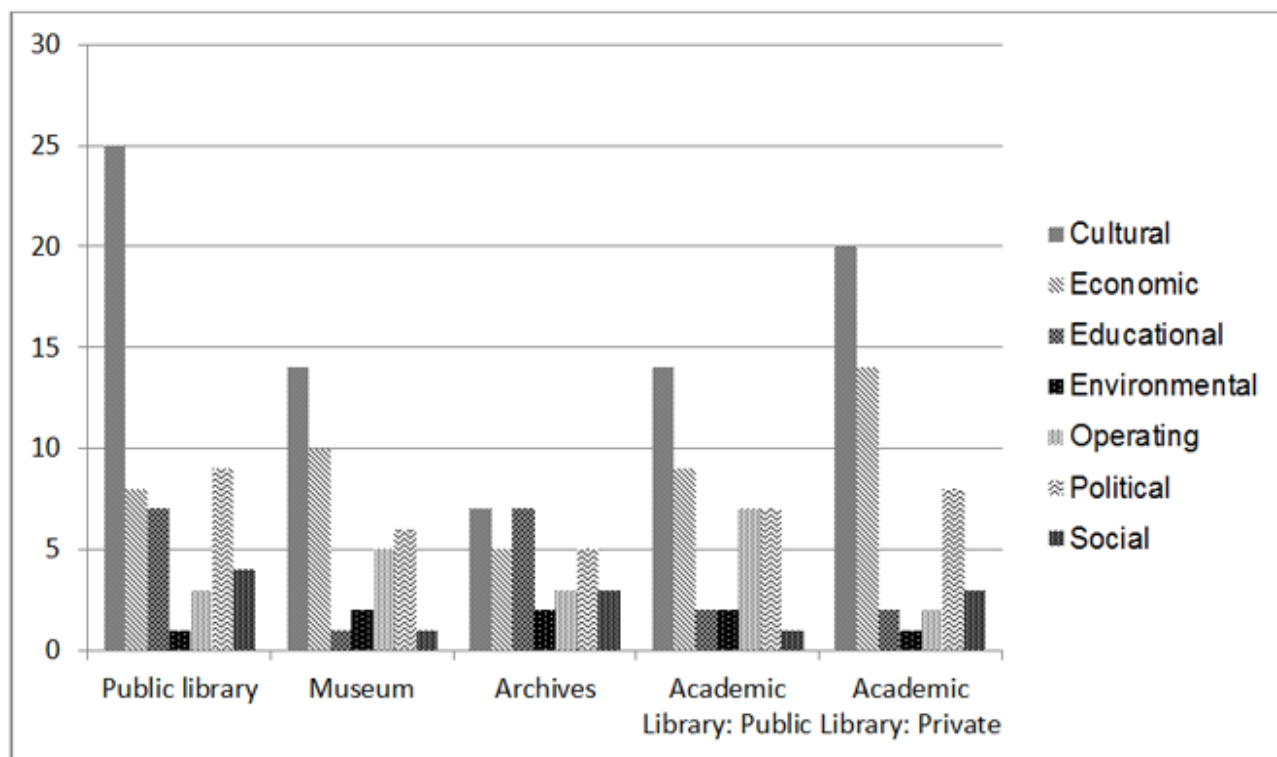## Measuring the Impact of Preserving Digital Assets: Partner Interviews

Figure 1. Areas of impact by institution type

16% of the value statements fell into political impact area, which included aspects such as greater visibility and prestige value. Partners also noted comments from visiting patrons about materials they had first seen online. One Portal partner noted that any action they took that increased the visibility of their name had a value for marketing and increasing awareness of their existence.

Operational Impact: "…the time it would've taken, from the staff time to set up the appointment with them, spending the time copying, scanning, finding, the item… the cost is really priceless for us. We have two people managing everything so time alone is invaluable but definitely in material costs as well. We'll just point them … directly to a link, send them and there it is. And that takes (snaps) that long."
  – Museum Director

Fully 9% of the total value statements fell into the operational impact area, which seemed to have a bigger impact in smaller institutions. The fewer personnel a partner had, the more savings in staff time accrued from having easy accessibility to their materials – both from an informational and operational perspective.

Educational Impact: "It's really a good time to be doing this kind of research, whether it be historical, genealogical, or whatever because there's this back and forth of the traditional way of doing this scholarship and the other, of all of this other high-tech stuff, which really isn't that high-tech anymore. You can walk a patron through learning how to use these resources and it's not really that difficult and the access is just…people are doing dances in the aisles on finding what they do find. It's a really exciting, exciting time."
  – Public Librarian

Nine percent of the value statements fell into the educational impact area. Several partners noted dramatic increases in the usage of their digital materials in publications, which they attributed to access and discoverability of their materials. One anecdote involved an historical society in another state which published an article in their journal about a letter collection they encountered through the Portal that included a correspondent from their area. The partner felt this connection would not have been made without the easy access digitization provides.

Social Impact: "We also have been exploring how to use things like Facebook and other kinds of, I have a feeling there are folks on our campus that our bloggers …"
  -- Academic Library Director, Private Institution

Findings revealed that few partners had delved deeply into social media channels with the exception of the museum partner. Some partners were not even aware if their online materials were being discovered and shared by online patrons through social media. Comments in this area were 5% of total value statements.

Environmental Impact: "People can find things easier, they can use the information more easily than going through the old ways of knowing who's got what and traveling to get there."
  – Academic Librarian, Public Institution

Four percent of the value statements fell into the environmental impact area, and every comment noted the value that accrues to patrons when they do not have to travel to see and use cultural heritage materials.

Challenges Identified: "In order to go beyond the normal type of statistics that are generated from the use of materials in archives and collections, we need to look at different ways of how people use things."
  -- Archivist

Challenges noted by Portal partners included a lack of overall knowledge in effective ways to assess the value of their online materials. Several also noted that while usage is a valuable metric, it doesn't tell you how items are actually being used and the level of value realized by patrons who discover them.

Usage Statistics
Although its value is generally acknowledged to have limits, usage remains one indicator of economic impact that is important to content providers. Aggregate data for all objects within the Portal system indicate that usage of Portal objects has continued to grow.
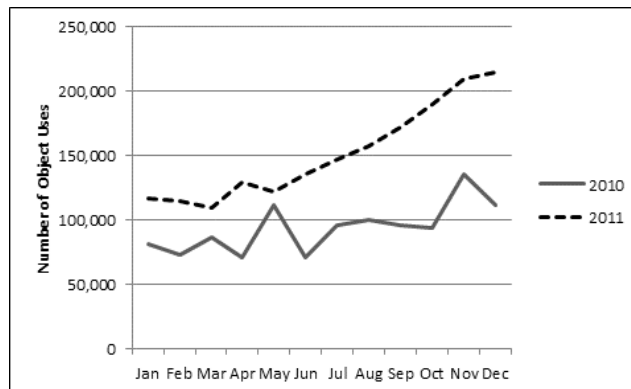

Figure 2. Aggregate system usage for The Portal to Texas History

There was a 61% increase in usage for the two full years (2010 and 2011) for which comparable usage statistics were available (Figure 2). Monthly usage during this period ranged from 70,888 "uses" in June of 2010 to 215,020 "uses" in December of 2011.

Impact on Users
 "I would like to think that the impact from greater access would help get people interested at maybe a younger age about the importance of history. … you can see … how things change, would improve knowledge. Hopefully a bit more intelligent society? That would be my dream."
  -- Museum Director

Scholarly Researchers: Citation Analysis
The citation study measured the impact of digitizing resources on scholarly researchers, and extracted citation data from publications in Google Books, Google Scholar, JSTOR, and

Academic Search Premier (N = 43). Analysis of citations revealed several trends.

- Not surprisingly, 40% of the citations resulted from historical publications.
- One surprise was the number of citations (35%) that occurred in the literature for Information Sciences.
- Citations spanned several other disciplines, including legal, music, biology, geography, tourism, and fiction.
- Publication formats included books (47%), journal articles (40%), theses (5%), and dissertations (9%).
- The number of citations per year from 2003 through 2011 increased 110%, perhaps indicating a relationship between the quantity of materials available on The Portal to Texas Hisotry (an increase of 3800% [8]) and the usefulness of the assets for scholarship.

Portal Users: Comment Log Analysis

Of the 2,422 impact areas and other categories agreed upon by the researchers in their analysis of user-submitted comments, 196 were designated as "not applicable" to the study. These were often notices of items for sale or requests for the physical location of historical artifacts. The remaining comments were primarily designated to "other categories", that is, not to areas of impact per se (N = 2,008) (Figure 3).
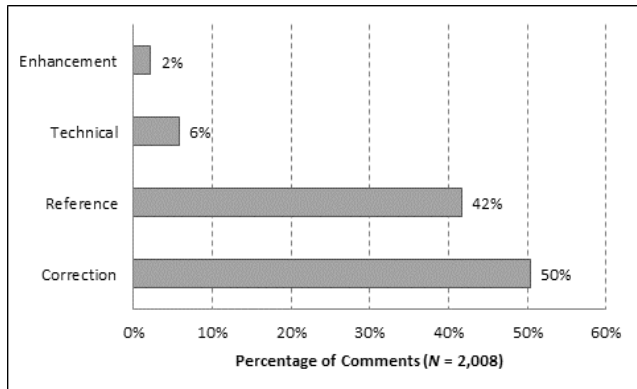

Figure 3. Indicators of dimensions other than impact

Most of these were in the 'Corrections' category, which included three indicators: typographic errors, image issues, or metadata additions. By far most of the comments pertained to metadata additions, and often for photographs. Users identified people, organizations, events, and locations.

Of particular interest to this study are those comments that indicated the impact of the Portal in the areas within the common framework (N = 218). Three of the seven impact areas were identified in the analysis: social, cultural, and educational (Figure 4). The greatest area of impact that emerged was cultural.
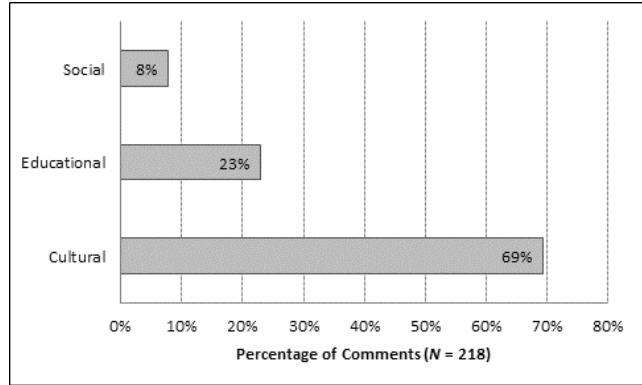

Figure 4. Areas of impact from user feedback

Offers to contribute materials were a common indicator of cultural impact; however, most comments pertained to the "existence value" of the materials themselves, for which there were many expressions of thanks that they were being preserved and made publicly available. As an example, one user commented: "I am very glad to have found this site. I am interested in the events relating to the assassination of JFK and I live in London, UK, where it is very hard to find first hand material on the subject."

Closing

There is some indication, from both the interviews with content partners and the comments from users, that the primary impact of digitizing resources for both of these groups is cultural. They seem to think that preservation of resources is a good thing to do and that access to these resources is beneficial. Still largely unknown are "the differences for the better" that preservation of and access to digital resources have made in people's lives.

Consistent with a multiple methods approach, the findings from the early study activities will inform future research activities. These will include different methods, and in some cases wider audiences. In regard to the impact of digitizing their holdings on The Portal to Texas History, a survey of all content partners will be conducted to gain further insights into the findings from the interviews with key informants. Additionally, statistical analysis of long-term usage data for individual content partners will be conducted and compared to the aggregate usage data for the Portal system.

In regard to the impact of digitizing resources on users, both a general survey of users and a specific survey of scholars will be conducted. Participants in the general survey will be solicited from a wide population of Portal users. The intention of this survey is to document "the differences for the better" that digital resources from The Portal to Texas History have made in individuals' lives. It may be that more areas of impact from the common framework will be elicited, as compared to the three areas that emerged from the comment log analysis. Lastly, scholars identified in the citation analysis will be invited to participate in a survey. In a similar vein to the planned survey of users, scholars will be asked to share their experiences relative to the value they derived from access to the Portal's resources.

Building on the early findings reported in this paper, future investigations will hopefully provide additional evidence of the impact and value of digitizing resources from the perspectives of two groups of digital library stakeholders who are external to the UNT Libraries: content providers and users. The findings may help other academic libraries striving to develop more meaningful impact and value measurements for their digital collections.

References

[1]   S. Chowdhury, M Landoni, & F. Gibb. "Usablity and Impact of Digital Libraries: a Review," Online Information Review, 30, 6. (2006) pg. 671.
[2]   P. Hirtle. "The Impact of Digitization on Special Collections in Libraries," Libraries & the Cultural Record, 37(1), 43 (2002). pg. 43.
[3]   T.B. Hahn, "Impacts of Mass Digitization Projects on Libraries and Information Policy," Bulletin of the American Society for Information and Technology, 33(1), 24 (2006). pg. 24.
[4]   A. Isfandyari-Moghaddam, "Digital Libraries in the Mirror of the Literature: Issues and Considerations," The Electronic Library, 26, 6. (2008)
[5]   S. Tanner, Inspiring Research, Inspiring Scholarship: The Value and Benefits of Digitized Resources for Learning, Teaching, Research and Enjoyment, Proc. Archiving 2011, pg. 81. (2011).
[6]   W. Edgar, "Corporate Library Impact, Part 1: A Theoretical Approach," Library Quarterly, 74, 2. (2004).
[7]   M. Oakleaf, Value of Academic Libraries: A Comprehensive Research Review and Report (ACRL, Chicago, IL,2010) pg. 87.
[8]   Portal to Texas History, Statistics for The Portal to Texas History: Items added between January 2004 and March 2012. Retrieved March 24, 2012 from http://texashistory.unt.edu/stats/.

Author Biography

Kathleen Murray received her PhD in information science from the University of North Texas (UNT). She is a postdoctoral research associate at the UNT Libraries. Her work primarily involves user studies in the areas of digital libraries and Web archives.

Dreanna Belden received her MS in information science from the University of North Texas (UNT). Subsequently she served as program director for The Portal to Texas History with responsibility for content partner recruitment and relationships, as well as external development. She is currently Assistant Dean for External Relations for UNT Libraries.