

Appraising our Digital Investment: Sustainability of Digitized Special Collections in ARL Libraries

A Report from Ithaka S+R and the Association of Research Libraries

Nancy L. Maron, Ithaka S+R

Sarah Pickle, Ithaka S+R

February 2013



Appraising our Digital Investment:
Sustainability of Digitized Special Collections in ARL Libraries

February 2013

Report Prepared for the Association of Research Libraries by

Nancy L. Maron, Ithaka S+R

Sarah Pickle, Ithaka S+R

For more information, please visit:

<http://www.arl.org/>

<http://www.sr.ithaka.org/>

Published by the Association of Research Libraries (Washington, DC) and Ithaka S+R (New York, NY)



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
KEY FINDINGS.....	2
STRATEGIES FOR ONGOING SUPPORT	3
INTRODUCTION	4
METHODOLOGY OVERVIEW	7
A NOTE ON DEFINITIONS	8
KEY FINDINGS.....	10
SURVEY RESULTS	11
THE INSTITUTIONAL PERSPECTIVE OF DIGITIZED SPECIAL COLLECTIONS	11
DIGITIZED SPECIAL COLLECTIONS AT ARL INSTITUTIONS.....	14
HOW LIBRARIES ARE MANAGING THEIR DIGITIZED COLLECTIONS.....	16
1. Responsibility for Managing Digitized Special Collections	17
2. Expenditures	19
3. Funding Sources	21
4. Understanding and Engaging Audience	28
CONCLUSION: STRATEGIES FOR ONGOING SUPPORT	32
APPENDIX A—ADDITIONAL NOTES ON METHODOLOGY	36
APPENDIX B—OPEN-TEXT RESPONSES TO PART 3: SPECIFIC COLLECTIONS.....	41
APPENDIX C—ONLINE SUPPLEMENT: SURVEY INSTRUMENT AND DATA	49

EXECUTIVE SUMMARY

Special collections have long been a vital part of libraries, offering users access to rare or archival materials and creating unique value for their host institutions. Research libraries are well aware of the particular value that their special collections hold and are increasingly placing greater emphasis on efforts like digitization that can make them more accessible to their users.

While many research libraries have begun to digitize their collections and share best practices around the steps required to create digital content, much less is known about what happens post-launch. Building on previous research by Ithaka S+R that defined key aspects of sustainable digital content, *Appraising our Digital Investment: Sustainability of Digitized Special Collections in ARL Libraries* offers a first look at the practices, attitudes, costs, and revenues associated with caring for digitized special collections. The report shares results from a survey conducted on the sustainability of digitized special collections at ARL member institutions.

The survey, created by ARL and Ithaka S+R, in collaboration with an advisory board and ARL’s Transforming Special Collections in the Digital Age Working Group, included three separate sections, in order to probe institutional priorities and understand current practices across the whole library and at the individual project level. Constructed using the web-based SurveyMonkey platform, the survey was sent to directors of all ARL member institutions, who were asked to complete the first section and then recommend others to complete subsequent sections, if desired. The survey opened on March 26, 2012, and closed 10 weeks later, with a further extension until July 11, 2012, to ensure a high response rate and encourage completion.

Survey section	Sent to	Response rate
Institutional Perspective	Library directors and leaders, those responsible for guiding the institutional mission	89 respondents, 70%
Collections in the Aggregate	Someone with first-hand knowledge of the range of activities needed to create and support all the digitized special collections at the library	70 respondents, 55%
Specific Collections	Someone who could report on one specific collection	50 respondents, 39%

In conducting the data analysis, we confirmed that there was no response bias and that the survey respondents were representative of all ARL member institutions by comparing the total and materials expenditures for those who responded with the expenditures of all ARL institutions. We also scanned the data for outliers and inconsistent responses and followed up with respondents to confirm or adjust unclear answers; where data could not be confirmed or corrected and where survey responses were more than 50% incomplete, we chose to omit those responses.

KEY FINDINGS

- **Most library leaders feel that digitized special collections are critical to the libraries' future, but few feel their institutions' investments in updates and upgrades are sufficient.** Over 80% agreed that digitized special collections are “critical to our current strategic direction” and yet almost a third feel they are under-investing in this area. Over half of respondents cited funding of this activity as their greatest sustainability concern.
- **Management is often dispersed across many departments, and no one department has emerged as the dominant locus of primary responsibility.** About 20% of respondents reported having created “digital” working groups or committees within the library; where there is an obvious “lead” it tends to be either the library IT department or Special Collections department. As many respondents told us when discussing their sustainability concerns about their digitized special collections, it is hard to coordinate the work of a committee or focus the work of an individual who has varied responsibilities so that they can sufficiently tend to the collections they manage and dedicate time to all of the activities that may be required to sustain them. This environment may make it difficult to budget or plan for growth.
- **Libraries are spending far more to create new resources than they are on maintaining and enhancing the ones they have already created.** Aggregate figures show the cost of ongoing support for all digitized special collections is just a fraction of the amount spent in any one year to create new ones, and the raw figures often represent small fractions of someone's time. This suggests a scenario where digitized collections, once created, are intended to essentially run without much active management, a situation that could ultimately hamper the ability of these institutions to sustain their projects and achieve the impact they desire.
- **While new digitization work is funded through a combination of internal and external sources, the expenditures associated with any ongoing support of collections are almost entirely covered by libraries' base budgets.** Libraries that began digitizing in 1999 or earlier show greater diversity in revenue sources for the creation of resources and more evidence of partnerships in supporting this work.
- **Support for open access is a main factor cited by many to explain a decision not to pursue revenue generation.** This is clearly a strongly held opinion by many respondents. Yet, most of those who do engage in revenue generating activities have found ways to do this that are open-access compatible. Others decline to pursue this avenue, feeling it will not be worth the investment.
- **Although the ability to offer greater access emerged as a key motivator for digitizing collections, investments in understanding the needs of the audience are quite low.** Forty-three percent of libraries are actively gathering some data about their digital audience, and this is most often in the form of web analytics and much less often involves what respondents describe as the more effective, qualitative approaches. In addition, while about half of respondents report conducting some form of regular outreach, only 17% of respondents say they measure its effectiveness.

STRATEGIES FOR ONGOING SUPPORT

Based on the findings included in the report, a set of recommendations is offered to guide project leaders in libraries as well as library directors, campus administrators, funders, and others who support the investment in and care for digitized special collections.

- Determine the desired impact of the digitized collections. If digitization is intended to preserve fragile originals, simply having created digital surrogates is likely sufficient. If the goal is to provide content to populate a digital library, there, too, perhaps most work is finished once the digitization phase is complete. But if the goal includes ensuring that the content has some measurable impact for its intended audience, then this will require an additional set of activities to understand user needs, to inform them about this material, and encourage engagement.
- Oversight for digitized special collections is a highly distributed activity, but this does not have to be a liability. It is crucial to clarify the roles of the individuals and departments involved, particularly as a project moves from its initial creation to an ongoing maintenance strategy. For those projects where ongoing attention is important, collaboration between the Special Collections department and the library IT team is critical to integrate the digital curation, preservation, and access activities of Special Collections into the institutions' overall digital infrastructure.
- If the intention is for the collection to have wide and/or deep impact among an audience or audiences, consider investing in qualitative research to learn more from those who are or could be using your collections. At very least, this will provide rich information about the value scholars and students place on the digitized collections. At the very best, it may also offer some useful revelations and suggest future sources of support.
- For those institutions whose digitized special collections demonstrate either high usage or strong, measurable impact among a narrow niche, consider what sources of financial support may be available beyond the core budget of the institution. There may be opportunities, such as sponsorship or licensing, to generate funds in ways that do not compromise the ability to keep content openly available, but, rather, offer some flexibility in budgeting for enhancements.

INTRODUCTION

Special collections have long been a vital part of libraries, offering users access to rare or archival materials and creating unique value for their host institutions.¹ Research libraries are well aware of the particular value that their special collections hold and are placing greater emphasis on efforts like digitization that can make them more accessible to their users.²

In recent years, significant investments in the digitization of these collections have been made both by the libraries and by government agencies and philanthropic organizations, including the Institute for Museum and Library Services (IMLS), the National Endowment for the Humanities (NEH), the National Science Foundation (NSF), The Andrew W. Mellon Foundation, the Canadian Heritage Information Network, and the Social Sciences and Humanities Research Council in Canada.

Along with their financial commitments, funders have become increasingly interested in thinking about the ongoing impact of the content they help to create in order to “preserve and provide access to materials of broad cultural and scholarly significance.”³ IMLS includes “sustainability” as one of the evaluation criteria for its National Leadership Grants, and Mellon “encourag[es] grantees to address the sustainability of their initiatives.”⁴ The NEH requires those applying for Humanities Collections and Reference Resources grants to both “describe how the project’s results will be maintained and supported beyond the period of the grant” and “discuss plans for ensuring ongoing access to digital content generated through the project.”⁵

Although the community’s understanding of what sustainability means is not yet firm and will certainly vary by project, some elements are generally accepted. Best practices for the creation of digital copies and the steps needed to preserve them, as well as the legal concerns that can arise in this process are

¹ The ARL Working Group on Special Collections discussed this value in *Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections* (March 2009), <http://www.arl.org/bm~doc/scwg-report.pdf>. See also Christian Dupont and Elizabeth Yakel, “‘What’s So Special about Special Collections?’ Or, Assessing the Value Special Collections Bring to Academic Libraries”, in *Proceedings of the 2010 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment* (Washington, D.C.: ARL, 2011), 417-426.

² See the report issued by ARL, the Coalition for Networked Information, and the Scholarly Publishing and Academic Resources Coalition, *Research Library Issues 279: Special Collections and Archives in the Digital Age* (June 2012), <http://publications.arl.org/2ds247.pdf>; Institute for Museum and Library Services, *Status of Technology and Digitization in the Nation’s Museums and Libraries* (January 2006), http://www.imls.gov/assets/1/AssetManager/Technology_Digitization.pdf.

³ Andrew W. Mellon Foundation, *Scholarly Communication and Information Technology*, http://www.mellon.org/grant_programs/programs/scit.

⁴ Andrew W. Mellon Foundation, *Grant Programs: Overview*, http://www.mellon.org/grant_programs/programs.

⁵ National Endowment for the Humanities, *Humanities Collections and Reference Resources Grant Description* (July 2012), <http://www.neh.gov/files/grants/humanities-collections-reference-resources-july-19-2012.pdf>.

reasonably well defined.⁶ Recent studies in the United Kingdom have examined the many ways in which impact can be measured for digital resources in the academic and cultural sectors.⁷

Prior studies have explored the nature of special collections, asking about their size (number of items), content types, staff assignments, expenditures, funding sources, and preservation and conservation programs,⁸ and have investigated the practices that are applied in digitization, such as the general kinds of work special collections staff dedicate to their digitized resources and whether they have plans for large-scale digitization.⁹ Still largely unknown, however, is what happens to collections and how they are used after they have been digitized.

Given the new requirements of the digital world, how can we determine whether “the project’s benefits will continue beyond the grant period, either through ongoing support of project activities or products, websites, and development of institutional expertise and capacity, or through broad long-term access to project products”?¹⁰ Ithaka S+R has explored the key factors of sustainability through case studies that examined different types of digital content projects in order to see how their project leaders had thought about sustainability strategies to increase impact, while also identifying the ongoing sources of support needed to ensure that the impact was sustained. That work argued that sustainability was not just a question of the permanence of a digital file, but the presence of an engaged user base and management that allowed the resource to remain valuable to those users.¹¹

⁶ See the report issued by ARL, the Coalition for Networked Information, and the Scholarly Publishing and Academic Resources Coalition, *Research Library Issues 279: Special Collections and Archives in the Digital Age* (June 2012), <http://publications.arl.org/2ds247.pdf>.

⁷ See the Toolkit for the Impact of Digitised Scholarly Resources, <http://microsites.oii.ox.ac.uk/tidsr/>; Simon Tanner’s report *Measuring the Impact of Digital Resources: The Balanced Value Impact Model* (October 2012) <http://www.kdcs.kcl.ac.uk/innovation/impact.html>; and Eric T. Meyer’s *Splashes and Ripples: Synthesizing the Evidence on the Impact of Digital Resources* (May 2011), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1846535.

⁸ See Judith M. Pantich, *Special Collections in ARL Libraries* (2001), http://www.arl.org/bm~doc/spec_colls_in_arl.pdf.

⁹ Jackie M. Dooley and Katherine Luce, *Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives* (October 2010), <http://www.oclc.org/content/dam/research/publications/library/2010/2010-11.pdf>. Institutions participating in this OCLC survey belonged to ARL, the Canadian Association of Research Libraries, the Independent Research Libraries Association, the Oberlin Group, or the RLG Partnership. See also Association of Research Libraries, *Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections* (March 2009), <http://www.arl.org/bm~doc/scwg-report.pdf>. For a perspective outside of North America, OCLC Research has also recently released the report *Survey of Special Collections and Archives in the United Kingdom and Ireland* (2013),

<http://www.oclc.org/content/dam/research/publications/library/2013/2013-01.pdf>.

¹⁰ *IMLS National Leadership Grants for Libraries* (December 2012) http://www.imls.gov/applicants/2013_nlg_libraries_guidelines.aspx.

¹¹ As defined in Nancy L. Maron, et al., *Sustaining Digital Resources: An On-the-Ground View of Projects Today*, <http://www.sr.ithaka.org/research-publications/sustaining-digital-resources-ground-view-projects-today>. See also other Ithaka S+R reports, including Maron and Loy, *Revenue, Recession, Reliance: Revisiting the SCA/Ithaka Case Studies in Sustainability*, <http://www.sr.ithaka.org/research-publications/revenue-recession-reliance-revisiting-scaithaka-sr-case-studies-sustainability>.

Among the key success factors identified in that report were: the ability to understand the needs of users and stakeholders of the project; the identification of activities needed to support them; and the delegation of responsibility to manage costs and develop reliable sources of financial and non-financial support. In addition, the presence of an actively engaged manager was needed to set and accomplish these goals.

In 2012, Ithaka S+R and the Association of Research Libraries (ARL) launched a survey of ARL member institutions in order to capture information about the support of digitized special collections both during and after their creation. The aim of the survey was to identify broad trends in these activities and to pinpoint challenges that those in the field face when attempting to create and support digital content. We also asked about the impact these resources are having, and, by requesting data on expenditures and revenues, provide benchmarks that would be useful to others undertaking this work.

With these facets of sustainability in mind, the survey looked well beyond issues of technical preservation to consider the long-term management of digital files, the efforts underway to understand and engage with current and potential audiences, and the funds that make all of these activities possible. Questions asked about the costs for tasks such as scanning, project management, preservation, copyright clearance, usage analysis, and outreach. The latter two were probed more deeply in order to gather specific information about how libraries understand and connect with their users and whether their usage analysis and outreach projects were considered effective or helpful. Finally, in light of the significant costs of these collections and the increasingly strained budgets of many institutions, we asked whether institutions had attempted to generate revenue from their collections.

We hope that the findings provided here will benefit not only project leaders in libraries, but also library directors, campus administrators, and funders who support these digitized special collections and who, therefore, may seek to understand the factors and variables that characterize successful projects.

Many individuals contributed to the research and writing of this report, which would not have been possible without generous support from the Association of Research Libraries. We especially thank Charles B. Lowry and Judy Ruttenberg for their support and guidance throughout the project, Lee Anne George for her painstaking attention to developing the online questionnaire, and Karla Strieb for her contributions to the initial project. Members of Ithaka S+R contributed in many ways, from deploying the survey to analyzing data and reviewing early drafts of the paper, in particular, Deanna Marcum, Roger Schonfeld, Matthew Staiger, Jason Yun, and Sabine Zander. ARL's Transforming Special Collections in the Digital Age Working Group (henceforth the ARL Special Collections Working Group), chaired by Anne Kenney until December 2012, played a key role in vetting and testing the survey instrument and provided feedback on the analysis and presentation of findings. An early partnership with the Institute for Museum and Library Services—later split into a separate project and forthcoming in 2013—meant that the writing of the survey also benefitted from the advice of an IMLS advisory board, and we are very grateful to them for their help: Edward Galloway, Nancy Gwinn, Anne Kenney, Kim Sajet, Victoria Steele, and Beth Yakel.

Most important, we are deeply appreciative of the enthusiasm of the library deans and directors and their colleagues at ARL member libraries who responded to the survey. The information we asked them to share was not always easy to gather, but it is hoped that the greater library community will benefit from their efforts. Although many individuals graciously shared their time to support our work, the responsibility for the contents of this report rests with the authors.

METHODOLOGY OVERVIEW

This project began as an effort to gather data on the sustainability strategies that research libraries use to support the digitized special collections they create. An advisory board consisting of leaders in the research library world helped guide the creation of the survey instrument,¹² and in collaboration with the ARL Special Collections Working Group, they vetted and tested the questions developed by the Ithaka S+R research team and ARL staff. This process helped both to shape the questions and answer choices to best reflect the issues the library community cares most about and to refine the format of the questions so that they were as clear as possible for respondents. The survey was constructed by ARL using the web-based SurveyMonkey platform and was administered online.

While the research team and the advisers agreed that it was important to ask those responsible for guiding library policy (i.e., the library deans and directors) questions about institutional direction, it was also clear that these individuals were unlikely to respond to requests for greater detail about actual costs and activities related to managing digitized special collections. For that information, we expected they would pass along the survey to someone more closely engaged in that work. Because of the complexity of having multiple people enter information into one form in SurveyMonkey, the survey was built in three parts so that each could be answered by a different person, if desired.

The three sections of the survey included:

- **Institutional Perspective:** This part included questions for library directors, those responsible for guiding the institutional mission.
- **Collections in the Aggregate:** This section required more detailed reporting and was intended to be completed by someone with first-hand knowledge of the range of activities needed to create and support digitized special collections. We asked for information to be reported in the aggregate, taking into account all of the digitized special collections held by the library.
- **Specific Collections:** In this part, respondents were asked to report on *one* specific collection. While we asked about many of the same types of activities as in the “aggregate” section, this one provided us with a sense of the activities and investments required for the management of a specific collection.

The initial invitation email to participate in the survey was sent to directors of all ARL member institutions and signed by then-ARL Executive Director Charles B. Lowry. The letter introduced the survey and asked them to respond to the first part via a link embedded in the email. This introductory

¹² Because this project was originally part of a larger IMLS National Leadership Grant award to ARL and Ithaka S+R, the survey benefited from the guidance of the original advisory board.

email also included links to the other sections and PDFs of each document, so that directors could review the entire instrument.

At the end of each section, the respondent was asked to nominate the person who would be filling out the next part; the research team checked SurveyMonkey daily in order to see who had been nominated and sent that person an introductory email with the appropriate link and PDF.

The survey opened on March 26, 2012, and closed 10 weeks later. The research team followed up to ensure a high response rate and to encourage respondents who had not finished the survey to do so. In the end, the deadline was extended until July 11, 2012. The final response rates were:

- Institutional Perspective: 89 respondents, 70% response rate
- Collections in the Aggregate: 70 respondents, 55% response rate
- Specific Collections: 50 respondents, 39% response rate

In conducting the data analysis,¹³ we first confirmed that the survey respondents were representative of all ARL member institutions, and not just the largest ones, by comparing the total and materials expenditures for those who responded with the expenditures of all ARL institutions.¹⁴ We then scanned the data for outliers and inconsistent responses and followed up with respondents to confirm or adjust their answers; where data could not be confirmed or corrected (e.g., reporting 300,000 unique items, rather than a number of collections), we chose to omit those responses. Finally, responses were only included from those who completed at least half of their section of the survey, because we could not be sure whether those who did not make it halfway had simply decided not to answer the remainder or had intended to return to it (and perhaps revise it) later.

A NOTE ON DEFINITIONS

This survey aimed to establish some benchmarks for activities, costs, and revenues concerning the creation and ongoing support of digitized special collections. While some terms have developed accepted definitions within the community of research libraries (“special collections,” for example, is defined by ARL “broadly to include distinctive material in all media and attendant library services”), agreed upon definitions for many other newer concepts are still emerging. The research team developed the following definition based on the survey *Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives* (see footnote 12) and in consultation with the ARL Executive Director and members of the advisory board and the ARL Special Collections Working Group. “Digitized special collections” were defined as:

Collections of rare or archival content that would ordinarily be kept in a secure special collections or archives facility that your institution has either digitized itself, chosen to have someone else digitize, or taken on the responsibility for managing (i.e., a collection created elsewhere that

¹³ An expanded discussion of the data preparation and analysis is included in Appendix A: Additional Notes on Methodology.

¹⁴ Explanations of statements of statistical representativeness or significance throughout the report are located in Appendix A.

your institution now hosts and manages). This includes collections of content that are intentionally created or curated by your staff (e.g., the papers of a famous writer, documents from the civil rights movement). A collection could include multiple format types on a specific topic, or a single content type. Some collections may include some ‘born digital’ elements (e.g., a writer’s papers that include print manuscripts as well as email correspondence). If you consider it a ‘collection’, please report on it here. This definition excludes digitized general collections, digital resources that your institution licenses from outside providers (such as commercial databases), and collections composed entirely of born-digital content, such as general administrative records of your host institution. It also excludes ad hoc patron scanning request for specific documents.

As the concept of sustainability is less well-defined in the community, rather than ask institutions to self-report on whether or not their digitized special collections were “sustainable,” questions probed activities, costs, and revenues related to those activities that we have learned are critical to the long-term sustainability and impact of digital content collections.

Though survey questions were tested and revised to be as clear as possible, some questions asked for financial data that are not typically tracked. For example, most digitized collections are not budgeted separately, as this work is often undertaken by many people across many departments. To report on the full costs of the ongoing activities around a collection, respondents often had to go, pen in hand, to colleagues throughout their libraries and ask how much time and money was spent on this or that activity. Respondents were encouraged to determine costs to the extent that this was possible and were instructed that approximations would be acceptable.

Finally, while the survey included three sections, only the first two, reflecting the perspectives of the institution and the aggregate activity of digitizing special collections at the institution, should be considered broadly representative of all ARL institutions.¹⁵ The third section had respondents report on a specific digitized special collection so that we could better understand the trajectory of specific projects and their costs and revenues. Respondents were each asked to choose a project that met at least one of the following criteria: involving significant investment, making use of an interesting or innovative model, presenting significant sustainability challenges, or showing signs of significant impact. The specific projects reported on, therefore, are neither statistically representative of all projects at that institution nor of projects at ARL institutions more generally. Rather, they offer us a view of the sustainability challenges and strategies at the project level. Where we do cite data from this section on “specific collections” it is intended to provide illustration of broader points made elsewhere in the paper.¹⁶

¹⁵ Data from the first two sections of the survey are available online. See Appendix C.

¹⁶ Appendix B includes responses to open-text questions about the sustainability challenges these resources have faced.

KEY FINDINGS

The following are the major themes that emerged from the survey:

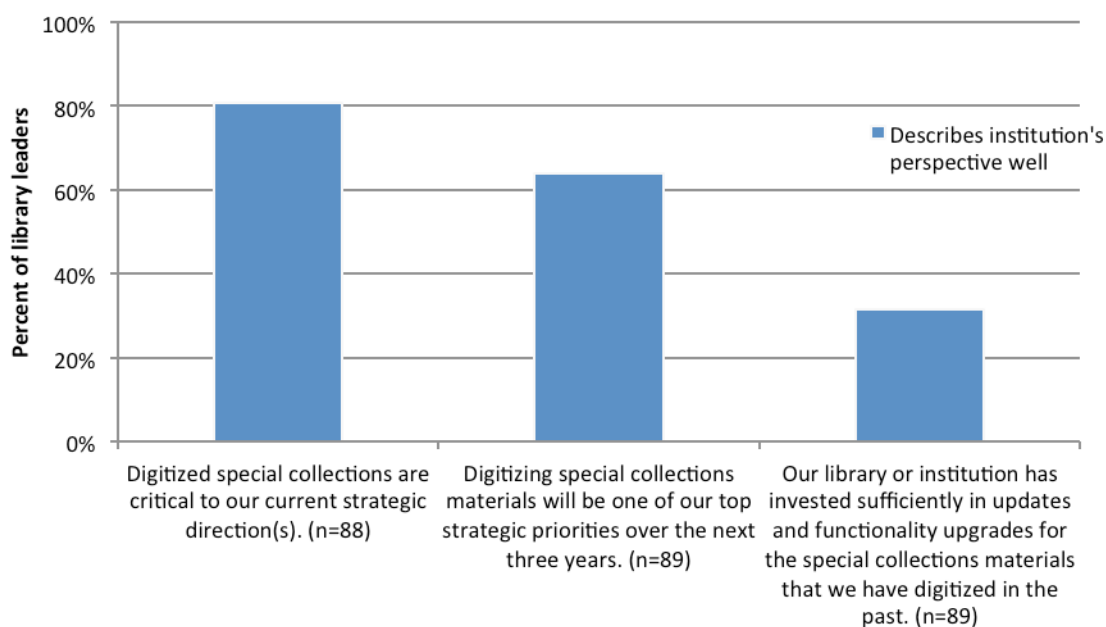
- **Most library leaders feel that digitized special collections are critical to the libraries' future, but few feel their institutions' investments in updates and upgrades are sufficient.** Over 80% agreed that digitized special collections are “critical to our current strategic direction” and yet almost a third feel they are under-investing in this area. Over half of respondents cited funding of this activity as their greatest sustainability concern.
- **Management is often dispersed across many departments, and no one department has emerged as the dominant locus of primary responsibility.** About 20% of respondents reported having created “digital” working groups or committees within the library; where there is an obvious “lead” it tends to be either the library IT department or Special Collections department. As many respondents told us when discussing their sustainability concerns about their digitized special collections, it is hard to coordinate the work of a committee or focus the work of an individual who has varied responsibilities so that they can sufficiently tend to the collections they manage and dedicate time to all of the activities that may be required to sustain them. This environment may make it difficult to budget or plan for growth.
- **Libraries are spending far more to create new resources than they are on maintaining and enhancing the ones they have already created.** Aggregate figures show the cost of ongoing support for all digitized special collections is just a fraction of the amount spent in any one year to create new ones, and the raw figures often represent small fractions of someone's time. This suggests a scenario where digitized collections, once created, are intended to essentially run without much active management, a situation that could ultimately hamper the ability of these institutions to sustain their projects and achieve the impact they desire.
- **While new digitization work is funded through a combination of internal and external sources, the expenditures associated with any ongoing support of collections are almost entirely covered by libraries' base budgets.** Libraries that began digitizing in 1999 or earlier show greater diversity in revenue sources for the creation of resources and more evidence of partnerships in supporting this work.
- **Support for open access is a main factor cited by many to explain a decision not to pursue revenue generation.** This is clearly a strongly held opinion by many respondents. Yet, most of those who do engage in revenue generating activities have found ways to do this that are open-access compatible. Others decline to pursue this avenue, feeling it will not be worth the investment.
- **Although the ability to offer greater access emerged as a key motivator for digitizing collections, investments in understanding the needs of the audience are quite low.** Forty-three percent of libraries are actively gathering some data about their digital audience, and this is most often in the form of web analytics and much less often involves what respondents describe as the more effective, qualitative approaches. In addition, while about half of respondents report conducting some form of regular outreach, only 17% of respondents say they measure its effectiveness.

SURVEY RESULTS

THE INSTITUTIONAL PERSPECTIVE OF DIGITIZED SPECIAL COLLECTIONS

In the first part of the survey, library directors, or others¹⁷ in a position to comment on the overall priorities of the institution, were asked about the role of digitized special collections, and most reported a significant place for digitized special collections in the future of their institutions.¹⁸ Over 80% agreed that digitized special collections are “critical to our current strategic direction(s),” and 64% agreed with the statement “Digitizing special collections materials will be one of our top strategic priorities over the next three years.”¹⁹ However, only 31% felt that they had invested sufficiently in updates and functionality upgrades for their digitized special collections (see figure 1).

FIGURE 1 How well does each of the following statements describe your institution’s perspective on digitized special collections?



¹⁷ Over 78% of those who responded to this section were directors of their institutions.

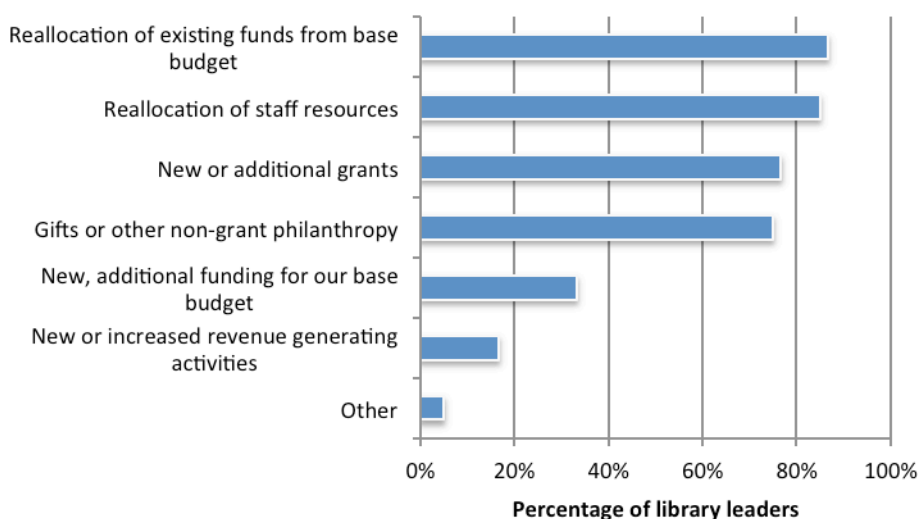
¹⁸ Perhaps because they serve as the primary public faces and advocates for their units within the university, library directors tend to value all the services their units provide. In the most recent edition of Ithaka S+R’s survey of library administrators, we asked library directors to rate the importance of six specific functions of the library (e.g., gateway, teaching facilitator, research supporter). Over 65%—and in most cases, more than 80%—of respondents indicated that all six roles are “very important” both now and will be “very important” five years from now. See Roger C. Schonfeld and Matthew P. Long, *Library Survey 2010: Insights from U.S. Academic Library Directors* (April 2011), <http://www.sr.ithaka.org/research-publications/library-survey-2010>.

¹⁹ These questions used a ten-point Likert scale for responses. Library leaders were asked to say how well a given statement described their institution’s perspective, where one equals “Does not describe our perspective at all” and 10 equals “Describes our perspective extremely well.” For the purpose of analysis and data presentation, the responses were banded into three groups, where answers ranging from 8–10 qualify as “describes institution’s perspective well.”

At the same time, approximately two-thirds of library leaders anticipate that over the next three years spending will increase either somewhat or substantially for both the up-front and the ongoing costs associated with these collections.

Where will the funding to cover these higher expenses come from? Over 80% of respondents who expect increases up-front reported that the funding to pay for these increases will come primarily from the reallocation of existing funds from the base budget and from the reallocation of staff resources. About 75% also anticipate receiving funding from grants and from gifts. Thirty-three percent expect this additional support to come from the institution’s base budget, and 17% identified new or increased revenue generating activities as a source of new funding.

Figure 2 Expected funding sources for increased spending for the creation of resources over the next three years (n=60)²⁰



As library directors acknowledge the significance of digitized special collections to their overall mission and seek new ways to invest in this area, identifying sources of support for this activity will not always be easy. When asked to describe “up to three of the biggest challenges to the long-term maintenance, enhancement, and preservation” of their institutions’ digitized special collections (see figure 3), 54% cited costs and securing funds to cover them:

- “Sufficient staff, money, staff. (Yes, that’s a deliberate duplication.)”
- “Finding the funds to support projects and achieve a high level of productivity.”
- “Limited funds for long-term management of special collections including archiving and delivery through innovative and up-to-date web portals.”

²⁰ “Other” includes spendable endowment income, capital funds from a new campus, and additional funds from the materials budget.

- “Funding for massive digital storage arrays and for the archival software support services needed to curate digital files.”
- “Long-term funding to continue to refresh the technology.”
- “Adequate funding to increase digitizing efforts, upgrade hardware and software and enhance metadata.”
- “Reallocating sufficient funds in a time of budgetary distress for medium to long-term curation.”
- “Economic downturn has made it more difficult to obtain grant and donor funding. [The] state-supported multi-institutional organization that should support a robust digital platform has not met our needs for specific features or scale, making it necessary to develop and fund this internally.”
- “Need for a consistent and growing source of funding to underwrite the costs of staff, etc., to increase digital collections.”

Figure 3 Biggest challenges to the long-term maintenance, enhancement, and preservation of your institution’s digitized special collections (n=83)²¹



In addition to funding, library leaders cited other concerns, including storage and preservation, the creation of standards for them, as well as thorny issues surrounding the management of staff time and expertise. On the latter point, one leader noted that “Our biggest challenge will be the permanent assignment of these responsibilities,” while another remarked upon “lack of expertise among special collections’ staff [who are] dealing with digital material.” Others cited the structural difficulty of

²¹ This was an open-text question, and respondents were asked to offer up to three answers. Responses have been grouped by common categories. Those remaining uncategorized responses (shown as “other” in figure 3) include: competing digital priorities, keeping up-to-date with user demand, diversity of content types and formats, big data, file checking (for redundancy, corruption), and digitizing materials before they deteriorate.

coordinating a project across a decentralized university campus and of preventing their resources from going stale; one expressed concern about finding “long-term funding to continue to refresh the technology.” Library leaders also anticipate challenges with understanding “how scholars might want to use the data in new ways and [how to ensure that] it is available to them in new forms of digital scholarship,” and they hope to be able to “[build] upon the strengths of our work with this collection” by exploiting social media and other outlets for outreach.

From the responses of ARL library leaders, it is clear that, although the digitized special collections at their institutions are highly valued, many questions—from funding and staffing to technology and institutional culture—remain about the how the work will be funded in the future.

DIGITIZED SPECIAL COLLECTIONS AT ARL INSTITUTIONS

In part one of the survey, library directors shared their views of the broad institutional directions and their concerns about the future of their institutions. The second part of the survey, however, offered further insight about the activities, funding, and costs concerning digitized special collections, as well as the libraries’ reasons for creating them in the first place.

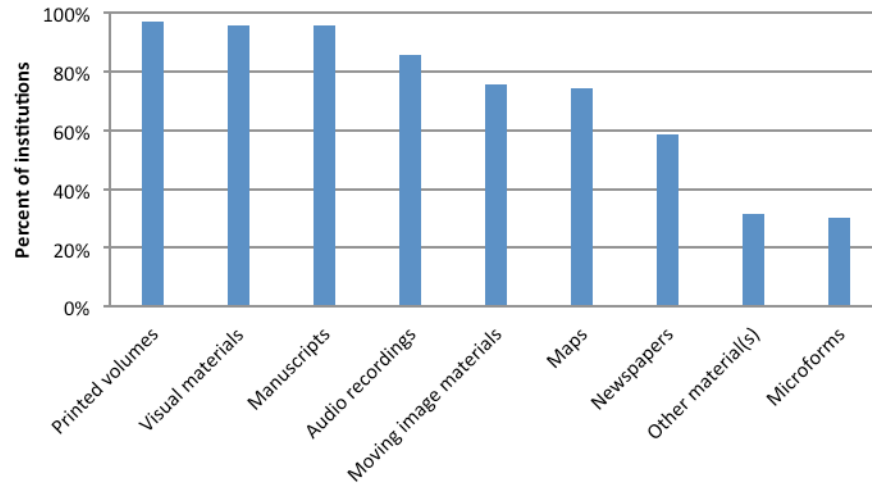
Over half of the ARL member libraries that responded to the survey have been digitizing special collections since 1999 or earlier. The earliest year recorded was 1988, and the most recent was 2009, with no apparent pattern of differences between when public and private institutions began this activity.

These institutions have digitized a wide range of materials (see figure 4). Nearly all have digitized printed volumes, visual materials, and manuscripts. Eighty-six percent also reported having digitized audio recordings, 76% moving image materials, and 74% maps. The variety of materials libraries are working with is quite broad. The median number of collections respondents reported was 30 with a range of three to 300 collections (see figure 5).²²

²² The median number of collections held by public institutions (35) was greater than the median number held by private institutions (25). One possible explanation for this gap might be that some public institutions might have more special collections materials from their states, and thus they might feel more of a duty to make those materials available to the greatest number of people. They may also be under greater budget constraints and see digitization as a cost-saving measure. Another explanation might simply be that different institutions have different methods for counting their collections. For example, one respondent, when asked for the approximate number of digitized special collections managed by her library, provided the number of “bibliographic items”; another shared how many “unique items” his institution has.

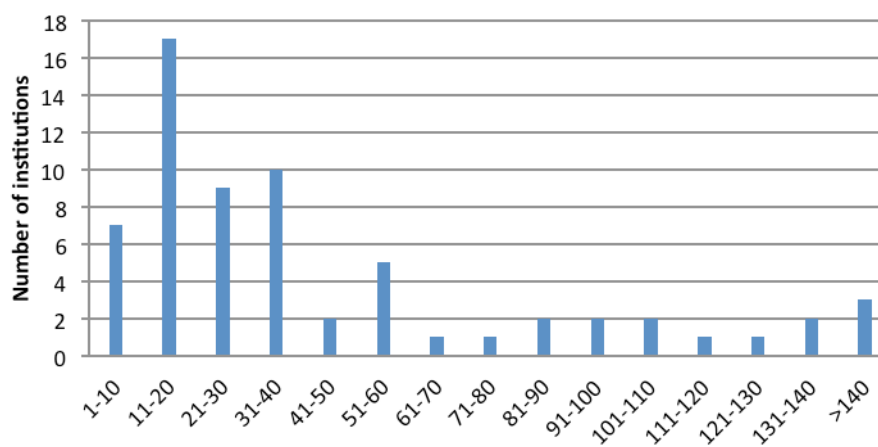
Figure 4

Content types represented
within the digitized special collections (n=70)



Other write-in responses to this question included:

- Born-digital files
- Artifacts
- Blueprints
- Architectural drawings
- Research data sets
- Advertising cards
- Cut-outs
- Charms
- Printed broadsides
- Correspondence
- Museum and scientific specimens
- Woodblocks
- Sheet music, music scores
- Study guides
- Pamphlet ephemera
- Scrapbooks
- University archives
- 3-d representations
- Digital humanities projects and exhibitions

Figure 5 Number of digitized special collections (n=65)

The digitized special collections that libraries hold come from a variety of sources. Almost all (99%) have themselves digitized some of their physical special collections; the other 1% described their digitized collections as consisting of born-digital materials or materials acquired from third-party vendors.²³ Just over 50% of respondents report hosting digitized collections created by other entities outside of their unit or institution. Almost one third (28%) receive copies of files created by publishers and vendors to whom they license special collections content for digitization purposes, and about the same percentage creates or acquires digitized collections through other processes, most often through collaborations with other institutions or as born-digital materials.

When asked to identify the factors that drove them to digitize their special collections, respondents named several, reflecting the many influences that spur this activity. Sixty-four percent reported that “collections strategy, based on prioritizing our strongest research areas,” was a highly motivating factor; 59% indicated that “user demand for the physical collection,” was very motivating; 54% chose “preservation, in order to protect fragile originals,” and 55% selected “donor relations or contributions.”

HOW LIBRARIES ARE MANAGING THEIR DIGITIZED COLLECTIONS

In order to understand how libraries are creating and supporting digitized special collections, the survey probed four main areas:

- (1) *Responsibility* for digitized special collections—who is involved in their care and where they sit within the organizational structures of their libraries.
- (2) The *costs* of creating and supporting these collections on an ongoing basis—e.g., the technical costs for keeping up these collections, the costs of staff time associated with

²³ Materials from vendors was listed by a respondent, even though the survey’s definition of “digitized special collections” asked respondents to exclude them.

creating, curating, and maintaining them, and the costs of staff time associated with acquiring and adding new digitized materials.

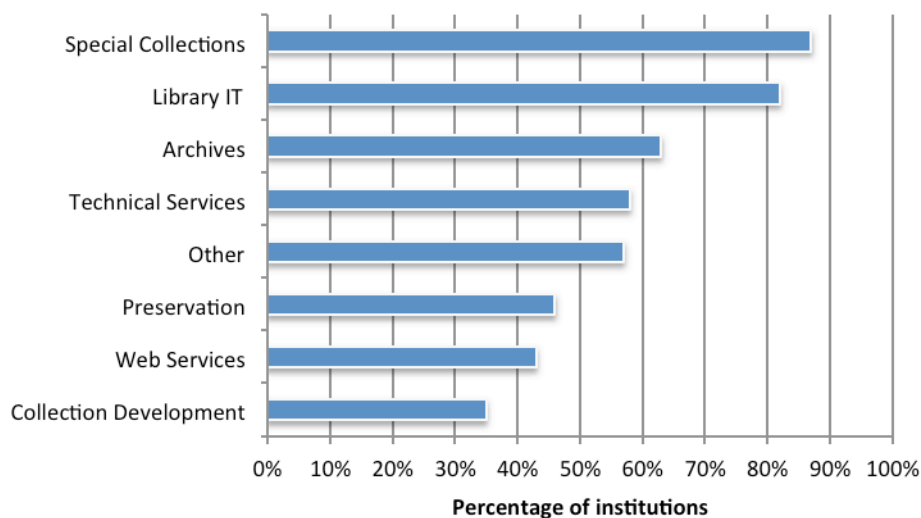
- (3) The *funding sources* libraries draw on to cover those costs—e.g., the base budget of the institution, grants, donations or individual philanthropy, or earned income.
- (4) How librarians seek to *understand their audiences and work with them*—through user needs analysis and other forms of research as well as and through outreach activities intended to promote the use of the collections.

The sections below discuss findings in each of these areas.

1. Responsibility for Managing Digitized Special Collections

The role of managing digitized special collections is typically distributed across several departments or groups within an institution (see figure 6). The departments or units that most often play a role are Special Collections (cited at 87% of institutions), library Information Technology (IT) (at 82% of institutions), and Archives (63% of institutions). On average, five departments help to support these collections, although three institutions reported drawing support from nine different departments. Three libraries assign this management to a single department.

Figure 6 Departments or units participating in the management or coordination of the ongoing support of digitized special collections²⁴ (n=89)



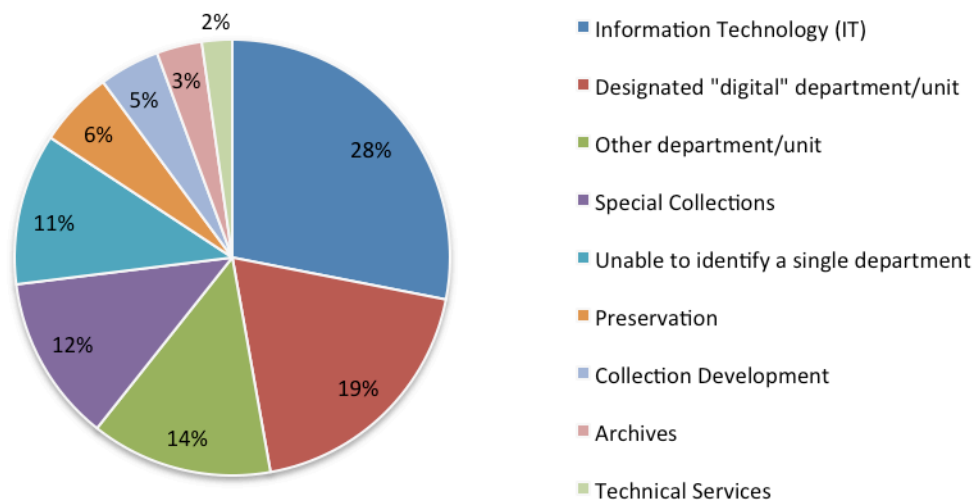
²⁴ “Other” includes administration as well as departments or committees that are unique to the respondents’ institutions, such as “Creation and Curation Services,” “Digital Services and Scholarly Communication,” and “Special Resources Portfolio.”

That said, when asked which group has *primary responsibility* for their digitized special collections, responses ran the gamut (see figure 7). The greatest number of respondents (28%) reported that the library IT department holds primary responsibility for these activities, while 12% named the Special Collections unit. Nineteen percent reported having created departments or units dedicated to “digital” work, the names of which include Digital Services and Shared Collections, Digital Library Services, Digital Scholarship & Programs, Digital Research & Publishing, Digital Initiatives, and Digital Collections.

“We currently have no single department focused on [the ongoing support of our digitized special collections], which has been identified as a serious flaw in our digitization activities.”
—A university librarian

But the message came across loud and clear that even if one department takes the lead, this is a shared set of activities, accomplished across many teams. Eleven percent noted that they were unable to identify a single, primary department. One respondent wrote that “We’re quite decentralized, so it’s a collaborative effort of special collections, Digital Content Creation, Preservation and IT.” Another noted that “We have several departments who together manage this environment. Our digitization efforts are not simple to explain [...]”

Figure 7 Department with *primary responsibility* for managing or coordinating the ongoing maintenance, enhancement, and preservation of the institution’s digitized special collections²⁵ (n=89)



²⁵ “Other” includes the one respondent who selected “web services” from the list of departments included in the survey and the eleven respondents who listed departments or committees that are unique to their institutions, as noted in the previous footnote.

2. Expenditures

The creation and ongoing care of digitized special collections can include a long list of activities, from acquiring, selecting, and digitizing materials to creating metadata, developing web interfaces, and undertaking outreach activities. In order to better understand the kinds of activities taking place, the costs they incur, and how they differ from the creation to the ongoing support phase, respondents were asked to report on the actual expenditures across all of their digitized special collections for the prior fiscal year. Respondents were encouraged to seek out cost data from colleagues in other units, as necessary, and to offer approximations if actual costs were too difficult to determine.

Most respondents (80%) were able to provide us with figures, although many acknowledged how challenging it was to gather them.²⁶ Among those unable to provide responses, several offered reasons, stating that these costs were too difficult to estimate, due either to the way their institutions typically account for these activities or to the distribution of the work: “It is impossible to know how much we are spending on the creation of digitized special collections,” said one. “We cannot begin to guess what our expenditures are for the creation and maintenance of digitized special collections as this work is being done across seven or more different units by a wide variety of staff, students and faculty and is not tracked,” reported another. Others offered additional explanation along with their responses:

- “We operate in a distributed model, so these activities take place in many departments and utilize staff and resources from around the library system. As a result, we are unable to provide the granular detail you are requesting.”
- “It is impossible to estimate our institution’s total expenditures for initial creation or ongoing maintenance, enhancement and preservation of our digitized collections.”
- “This is very difficult to quantify; it’s a small percentage of many people’s jobs. Some of it would be happening whether or not the collections were digital. While I appreciate the impetus behind this survey, I fear that any ongoing pressure to track these costs will result in either curtailing them or waste time that would be better spent doing them!”
- “I am sure you are aware that libraries generally don’t track staff time in ways that allow us to accurately estimate these costs. [...] I am grateful that this work is being done, but I will probably take the reports on cost estimates for some of these activities with a very large grain of salt.”
- “By the way, we’re totally guessing on questions 21–24. Hope that’s okay with you.”

With these caveats firmly in mind, the data provided below offer a first opportunity to observe what libraries are spending on the up-front creation and ongoing support for digitized special collections. Tables below display the mean and median, as well as the minimum and maximum figures for these

²⁶ Among those who provided answers to these questions, several put in placeholders in multiple fields (e.g., “1”) simply to signify that some amount of money had been spent on a given activity. Because it is impossible to determine why a respondent gave answers for some activities and not for others (did the activities cost nothing or were they not incurred?) and why some indicated “0” for some and left others blank, *the costs discussed in this report only represent institutions that have expenditures greater than \$0 for any given activity*. Because of this adjustment, the mean and median figures presented here may well be higher than they actually are. See Appendix A for further discussion of this decision.

activities, offering a sense of the overall costs of each of these activities and of their costs relative to one another. The stark differences in most means and medians underscore just how skewed the data are by some of the institutions (both public and private) who spent significantly more than their peers. Nevertheless, it is clear using either measure that the activities with the greatest costs were scanning, project management, metadata creation, and software development. On the lower end were usage analysis and user outreach and support, both of which are activities that are typically more common after a resource is launched.

Table 1 Up-front expenditures for digitized special collections (USD)

	Scanning	Project management	Copyright clearance	Metadata creation	Web design/software development	User outreach & support	Usage analysis	Preservation	Editorial
Mean	\$124,833	\$62,599	\$10,952	\$66,049	\$58,023	\$11,112	\$4,341	\$15,555	\$22,860
Median	\$48,500	\$36,768	\$3,000	\$35,000	\$30,000	\$3,000	\$4,550	\$7,510	\$8,817
Minimum	\$490	\$500	\$100	\$1,375	\$150	\$1	\$110	\$446	\$1
Maximum	\$932,000	\$452,145	\$70,000	\$425,000	\$330,000	\$50,000	\$12,000	\$100,000	\$239,000

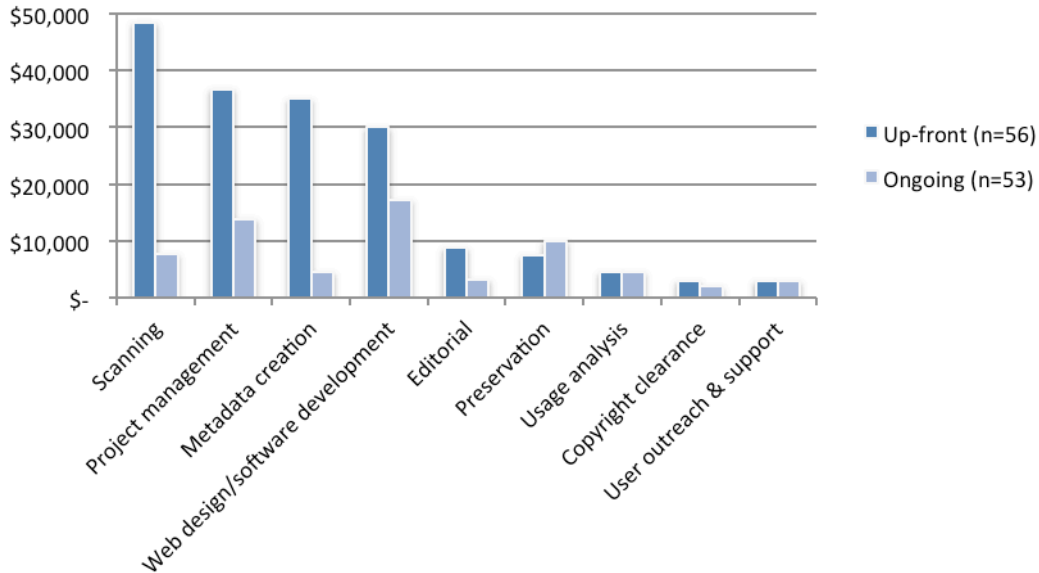
The money spent on these collections after they were created is presented in table 2. Once again, scanning, project management, and software development were among the highest expenditures, and preservation—an activity that must necessarily come after something has been created—joined them. Expenditures were low for usage analysis and for user outreach and support.

Table 2 Ongoing expenditures for digitized special collections (USD)

	Scanning	Project management	Copyright clearance	Metadata creation	Web design/software development	User outreach & support	Usage analysis	Preservation	Editorial
Mean	\$34,970	\$22,608	\$7,828	\$31,017	\$48,892	\$8,101	\$7,464	\$39,090	\$12,134
Median	\$7,645	\$13,923	\$2,000	\$4,520	\$17,259	\$2,900	\$4,550	\$10,000	\$3,209
Minimum	\$1	\$1	\$200	\$105	\$500	\$1	\$130	\$1	\$200
Maximum	\$285,000	\$140,000	\$70,000	\$630,000	\$958,594	\$53,000	\$71,000	\$732,000	\$135,000

These figures demonstrate that institutions are continuing to make significant investments in digitizing special collections and that these investments in new creation far surpass the level of investment to maintain and enhance all of the collections an institution has already digitized. The median institutional expenditures for new digitization came to \$218,787, while the median costs for the ongoing support of all existing digitized special collections came to \$61,500 (see figure 8). In general, institutions with older digitization programs (beginning in 1999 or earlier) did not spend money any differently than those with newer programs. The one exception to this is preservation, for which newer programs spent more than older programs at the creation stage (see Appendix A).

Figure 8 Median up-front and ongoing expenditures for collections in the aggregate (USD)



Certainly, some of the activities required to digitize special collections are costly, whether in staff time to create careful metadata or to scan fragile documents, so it is no surprise that up-front costs in some categories were quite high. Since respondents reported aggregate figures rather than costs for creating individual collections, it is difficult to know how many discrete projects are being discussed. If a library had major digitization efforts in development in a given year, that would result in higher totals, as well.

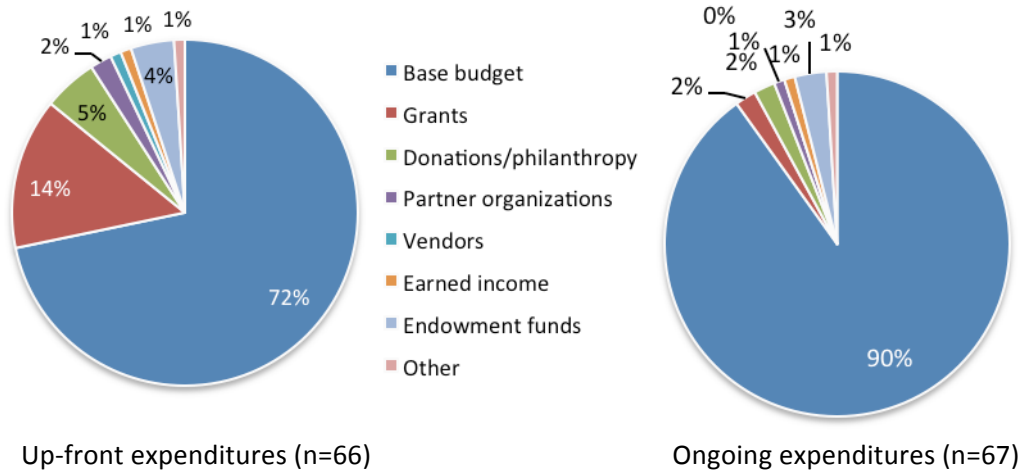
Even so, the raw figures reported for some of the ongoing activities suggest that minimal staff time is being devoted to them. The figures listed represent—just as do those listed for the up-front costs—the total cost of maintaining *all* the digitized collections the institution holds. Had we asked respondents to report on individual projects, the figures would likely be even lower.

3. Funding Sources

As ARL institutions continue to invest in the activities needed to create and support digitized special collections, where will this support come from? For new digitization projects, the lion’s share, 72%, comes from the base budgets of the libraries (see figure 9). The next largest share comes from grants, which account for 14% of these costs. The remainder is made up of a range of sources including donations, the work of partner organizations, and endowment funds.

Figure 9

Average percentage of funding for digitized special collections by source²⁷



Those institutions with longer experience with digitization show greater ability or willingness to secure initial external funding (see Appendix A). Libraries that began digitizing in or before 1999 received an average of 37% of their funding for the creation of their collections from external sources, whereas newer programs received an average of 19%. This difference is further amplified if one considers just those programs beginning in or before 1995 versus those who began digitizing later. Those that began digitizing in 1995 or earlier drew 44% of their funding from external sources, while those that started after 1995 only drew 26% from external sources. Here, the role of partnerships is significant: those programs starting in 1995 or earlier received an average of 9% from partner organizations, while those who started in 1995 or later only reported receiving 1% of their needed resources from partner organizations. Developing and nurturing partnerships takes time, and the older programs might have been able to take advantage of these opportunities before others began digitizing.

“It is often easier to acquire support for the creation of something new than it is to find support for the costs associated with maintenance, enhancement, and preservation. In a nutshell, funding is the issue.”

—A dean of a university library

Once a collection has been digitized, however, the sources of funding shift quite a bit. Respondents reported that 90% of the resources needed to support the digitized collections was supplied from the base budget of the institution. Just 10% came from other sources, including donations (2%), grants (2%), and the endowment (3%). None of the programs that began in 2000 or later reported that they received

²⁷ “Other” includes city and state operating budgets, federal work-study money, government-targeted funding, student technology fees, and Universal Service revenues.

help from partner organizations for the ongoing costs of supporting collections, and just four of the older programs did.

Although they depend less on the base budgets of their institutions for the support of their digitized special collections, the institutions that have been digitizing since at least 1999 have much larger budgets than their peers: on average, their total library expenditures are \$16 million higher and their materials expenditures are \$3 million higher than those libraries that began digitizing more recently. One possible explanation for this difference is that the wealthier institutions simply had the resources to begin digitizing earlier and thus also had more experience with or opportunity to experiment with seeking different funding sources.

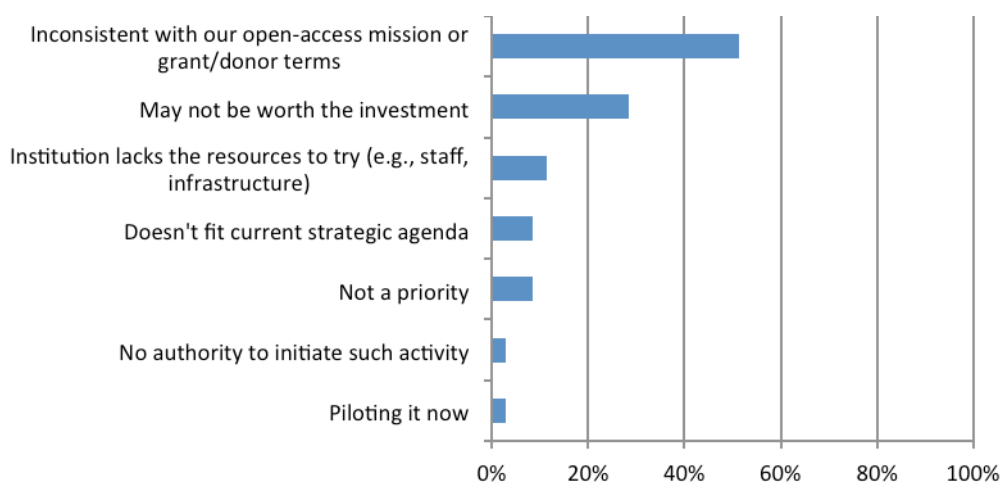
Revenue Generation

Beyond money raised through grants or from individual donors, research libraries only rarely generate revenues of their own to support their activities. Some, such as the Oxford University Libraries and the New York Public Library, have enterprise operations and gift shops, which provide at least some amount of earned income. But for most research libraries, general university budget funding, endowment income, and operating gifts are the principal sources of financial support. It should, therefore, be no surprise that the ongoing expenses associated with special collections are typically funded through base budgets rather than through other revenue sources. Nevertheless, many institutions have generated revenue from their digitized special collections, perhaps as a means to help support those resources or to contribute to the libraries' finances.

Respondents demonstrated a range of strongly held opinions about the suitability and efficacy of this activity, some of which may be due to lack of consensus around definitions of "open access" and "revenue generation." While about half of all respondents have experimented with some form of revenue generation, clear success stories are harder to find.

Just over half (51%) of the respondents report that their institutions have not tried any revenue generation. When asked to explain why their institutions had not attempted to generate revenue from their digitized special collections, many respondents offered multiple reasons. Figure 10 displays these responses grouped into common categories.

Figure 10 Reasons respondents gave for why their institutions have not attempted revenue generation from their digitized special collections (n=34)²⁸



The most common answer, offered by more than half of those who have not tried revenue generation, albeit with several variations, was that they felt revenue generation would be inconsistent with their open-access mission or grant/donor terms:

- “Our materials are open access.”
- “That is not consistent with our mission [or] with our ideals.”
- “We prefer open source access, rely on digitizing materials to which we have copyright (or PD), [and] consider it to be excellent outreach.”
- “We view digitization as a service to increase access. The access and profile is good PR and good for the University’s public image.”
- “We prefer cost recovery.”
- “As a state institution, we are not to generate revenue, just cost recovery.”
- “It is contrary to our mission and culture. The purpose for digitizing the collections is to ease discovery and use.”
- “Inconsistent with our educational mission and perhaps not well organized to do so.”
- “The library does not wish to put staff and resources towards creating a revenue model.”
- “The donor wishes the material to be freely accessible.”
- “Two of the three digitized collections are the result of federal grants which explicitly makes them freely available, and the third collection is a gift of a donor who wishes the collections to be freely available.”

These responses suggest certain assumptions: that revenue generation would be incompatible with a cost-recovery model and that revenue generation would be inconsistent with open access or otherwise

²⁸ Answers to this question were open-text and have been grouped by common themes. Some respondents offered more than one reason for not having attempted to generate revenue.

hinder discovery and access of the collection. Several statements indicate that, to some, revenue generation is viewed as entirely incompatible with open access.

Some clues to this discrepancy were evinced when survey results were previewed at the ARL Membership Meeting in October 2012. Some people equated the phrase “revenue generation” with selling goods or services at a profit, as opposed to simply offering them—regardless of pricing strategy—in exchange for money. Survey results and the open text answers summarized above also demonstrate a conflation of the idea of “revenue generation” with one particular revenue model that involves selling access to content for a fee.

There are, in fact, many options for generating revenue, either based on the value of the digitized content or the services developed around it, many of which permit free access to the core content. Much has been written in other sectors about the “freemium” models that permit free access to viewable files online or downloads of images at low resolution, while charging a fee for professional-grade versions of that content. Museums, in particular, have become quite adept at creating professional image licensing businesses while offering free content online. Open access is also fully compatible with advertising, as can be seen on many newspaper websites, corporate sponsorships, and the sale of products using images from collections (e.g., prints and postcards).²⁹

Indeed, 63% of those whose institutions *are* engaging in revenue generating activities agreed with the statement “revenue generation would be welcome *only* if it is compatible with open access for non-commercial purposes.” Among the models these libraries have tried are licensing and selling content to individuals and to other institutions, as well as sale of related services. Freely providing access to content, then, is a value shared widely by all.

Among those who have not experimented with revenue generation, 29% stated that they are not convinced that attempting to do so would be worth the investment. In the words of the respondents:

- “We lack the time, staff, [and] resources necessary to try a revenue model.”
- “Most[ly] due to a lack of staff resources to devote to the effort.”
- “We believe that administrative costs would be greater than the income earned. Effort required to generate a revenue model.”
- “In our system it would most likely cost more than it would generate.”
- “Time investment needed.”
- “Dubious about realizing a high enough return on the investment of staff time.”

It would certainly be important to understand more about how much time and effort programs have needed to devote in order to successfully develop and test their revenue models. This may point the way to help libraries evaluate their opportunities and to help them narrow down the choices they face.

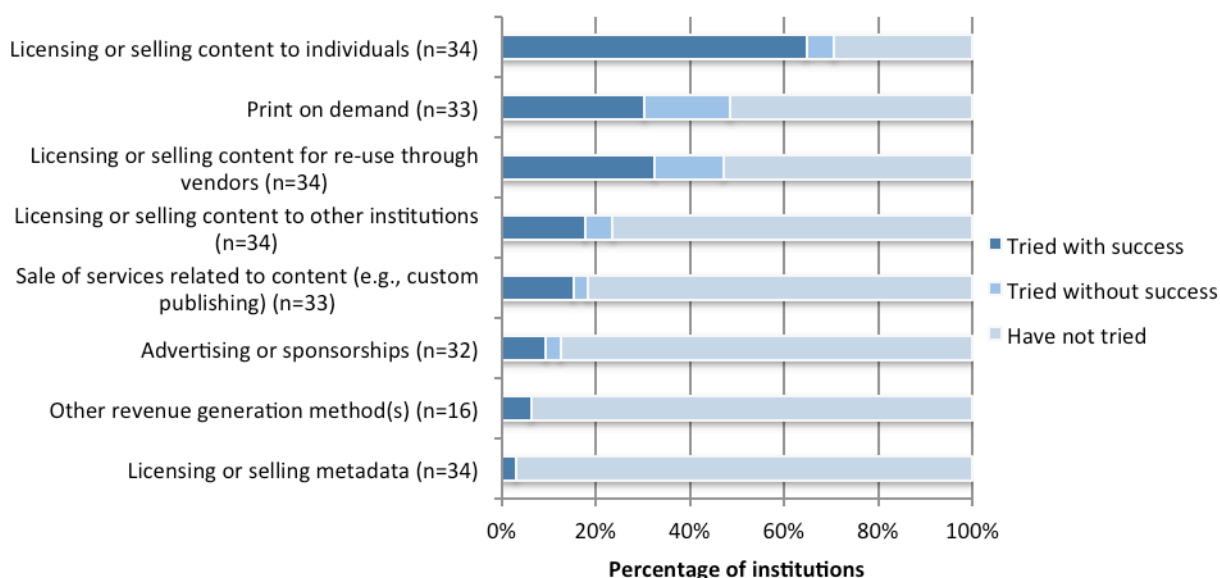
²⁹ The Ithaka S+R report on *Sustainability and Revenue Models for Online Academic Resources* (2008) outlines the range of types of revenue generating methods, including many compatible with open access. See <http://www.sr.ithaka.org/research-publications/sustainability-and-revenue-models-online-academic-resources>. A new edition of that report is forthcoming in 2013.

The fact that those who rate themselves as successful revenue generators (see figure 11) have all been digitizing for twelve years or more does suggest that, over time, the investment may pay off.

For those institutions—49% of all respondents—that have engaged in revenue generating activities, the most common method reported is licensing and selling content to individuals: 24 out of 34 institutions have tried it, and all but two respondents reported trying it “with success” (see figure 11).³⁰ Of the 16 libraries that have tried print on demand, 10 reported success, and of the 15 that have attempted licensing or selling content for re-use, two-thirds reported success.

Figure 11

Methods of revenue generation for the subset that has attempted it³¹



We also asked those whose institutions had generated revenue from licensing to tell us more about their arrangements. Around 80% indicated that they receive one-time payments, and around 40% received ongoing disbursements. The vast majority of these institutions (76%) received payments that are structured as flat fees. In addition, 33% received payments from royalties and 24% from revenue sharing. In terms of rights arrangements, 68% of institutions granted non-exclusive rights to the materials they license, and 41% granted exclusive rights for a designated duration. Only 5% granted non-exclusive rights with no limit on duration.

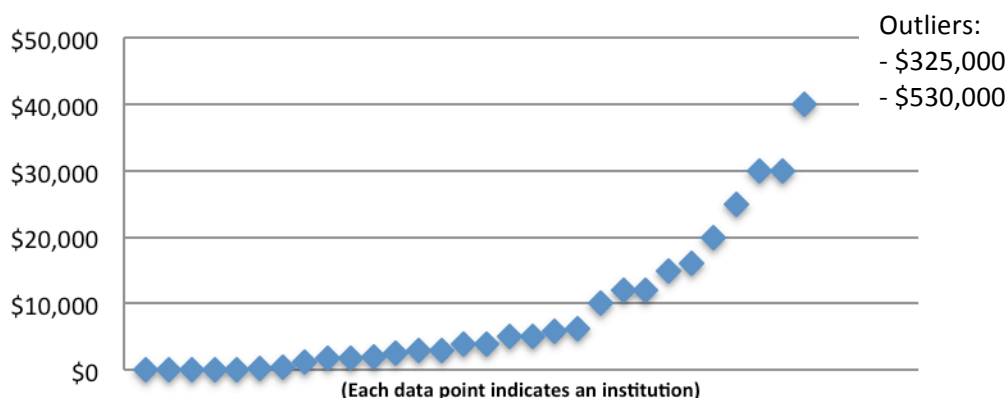
³⁰ “Success” is defined by the respondents themselves. Those who said that their institutions had tried revenue generation were asked to share more about the methods they had experimented with and whether or not they had tried them “with success” or “without success.”

³¹ The “other method” listed here was “licensing to third-party product developers” for products such as note cards and tea towels.

When these same respondents were asked to describe their most successful licensing arrangements, five mentioned the payments or royalties they receive through a vendor. One of them provided more detail: “Although we have made some successful arrangements with Hollywood, such as HBO and Disney, the most successful arrangements have been with museums and historical societies [...]” Another respondent explained that the library’s most successful arrangements are with Amazon.com, through which it sells audio files, and the university’s press, through which it sells digitized copies of “out of print/out of copyright/orphan books.”

The amount earned in the prior fiscal year by these institutions ranged from \$0, as reported by five libraries that described their efforts as “not successful,” to \$530,000, for a private institution with one especially strong revenue-generating collection. Excluding two institutions whose earnings were significantly higher than their peers’, all the others made \$40,000 or less, with an average of \$8,545 earned and median earnings of \$3,931 (see figure 12). This is not an especially high number, particularly when compared to the amount these collections cost to build and support. On average, the amount of revenue generated by each institution was about 21% of the total amount of money spent on the ongoing support of their digitized special collections in the prior year. The median for this ratio, however, was just 10%, and only two institutions earned more than half of what they spent in ongoing costs. While we do not know whether the revenue generated returns to the digitization program, the gains do seem modest when compared with the expenses.

Figure 12 Approximate revenue generated in prior year from digitized special collections, by institution (USD) (n=30)



Most institutions (76%) that have experimented with revenue generation reported trying at least two methods. However, experimentation is not always required if a sound method emerges early. While an institution that earned more than \$300,000 had tried four different methods and, in the respondent’s opinion, had been successful with them all, the institution that reported earning over \$500,000 had only tried—and been “successful” with—two. Thirty-four percent have tried two approaches, while 30% have tried three.

Some interesting results emerged from a closer look into which libraries have experimented with revenue generation.³² Sixty-three percent of those that began digitizing in 1999 or earlier have attempted revenue generation, while only 38% of those beginning in 2000 or later have attempted it, and those that have earned the most money have all been digitizing for some time. Additionally, the institutions that have tried to generate revenue tend to have more collections (mean of 63) than their peers (mean of 38).

We also found relationships between revenue generation and the perspective of library leaders. More of those who believed their institutions were described well by the statement that “digitized special collections are critical to our current strategic direction” belong to institutions that have tried revenue generation. On the other hand, more library leaders who answered that their libraries were described well by the statement that “The primary purpose of digitizing our special collections is to preserve or protect the physical objects” belong to institutions that have not tried revenue generation.

4. Understanding and Engaging Audience

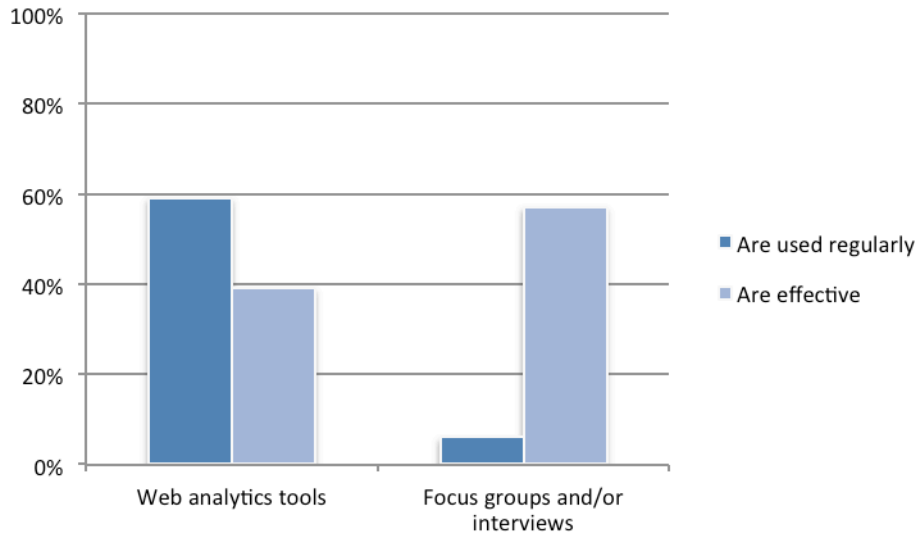
As noted earlier, 59% of ARL member libraries indicated that their decisions to digitize their special collections were highly motivated by user demand for the unique and distinctive content they hold. Given this emphasis on users, what steps are these libraries taking to understand this online audience and its needs?

In order to understand their audiences, 59% of institutions regularly use web analytics (e.g., Google Analytics) and 44% regularly examine user feedback via their websites. Far less regularly used are the assessment methods that take more time and money to develop, such as surveys (3%), focus groups and/or interviews (6%), and user feedback via social media (e.g., Facebook and Twitter) (24%).

Respondents were also asked just how effective they believed these methods to be. Despite the widespread usage of web analytics tools, fewer than 40% of respondents viewed them as being effective; however, focus groups and/or interviews are viewed as effective by 57% of respondents (see figure 13). Although the results of this survey cannot explain this discrepancy, it might simply be due to the greater cost and complexity of conducting focus groups and interviews.

³² See Appendix A for a discussion of the tests run to assess the statistical significance of these findings.

Figure 13 User needs assessment methods, by percentage of institutions (n=70)



In addition to any efforts directed at understanding audience, we wanted to know if libraries have sought to connect with their users and promote their digitized special collections, given that the impetus for developing these resources is, for so many of the institutions, to provide greater access to the materials.

Of the institutions surveyed, 43% report that they track or otherwise collect information about their users to determine the composition of their audiences. In this respect, libraries with older digitization programs behave no differently than those with newer programs. However, more of the libraries that do track users are led by library directors who believe that “our library or institution has invested sufficiently in updates and functionality upgrades for the special collections materials that we have digitized in the past” (see Appendix A).

“We need to develop additional tools to allow users to ‘open up’ the materials we have digitized and then feed back to us the enhancements they make or the visualization tools they created.”
—A library director

We asked respondents how often their institutions conduct specific outreach activities such as instructing classes or groups on methods for using the materials, communicating with researchers and students, promoting the collections on their websites, and creating finding aids. Respondents were asked to report whether they engaged in the activity regularly, occasionally, rarely, or never.

While almost every institution undertakes a variety of these approaches to some degree, two activities stood out from the rest. Nearly 80% of libraries reported “regularly” creating finding aids and

highlighting the digitized special collections on their public websites. By contrast, the next most common responses for activities used regularly were communicating with faculty and researchers (44%); instructing classes or groups in these methods (43%); and using social networking tools (41%). Other responses were even lower, including sending emails to known users, conducting outreach to new potential users, and conducting outreach to senior administrators on campus (see figure 14).

While creating finding aids and providing content on the website both require a great deal of effort and can be tremendously useful, they are fundamentally “passive” tactics, relying on the expectation that users will seek these materials out once they are made available. The other activities—those with much lower uptake among the responding libraries—are those requiring active outreach, i.e., actions that seek to identify, reach out to, and directly communicate the presence and value of collections for those who might use them.³³ As the web only continues to grow and

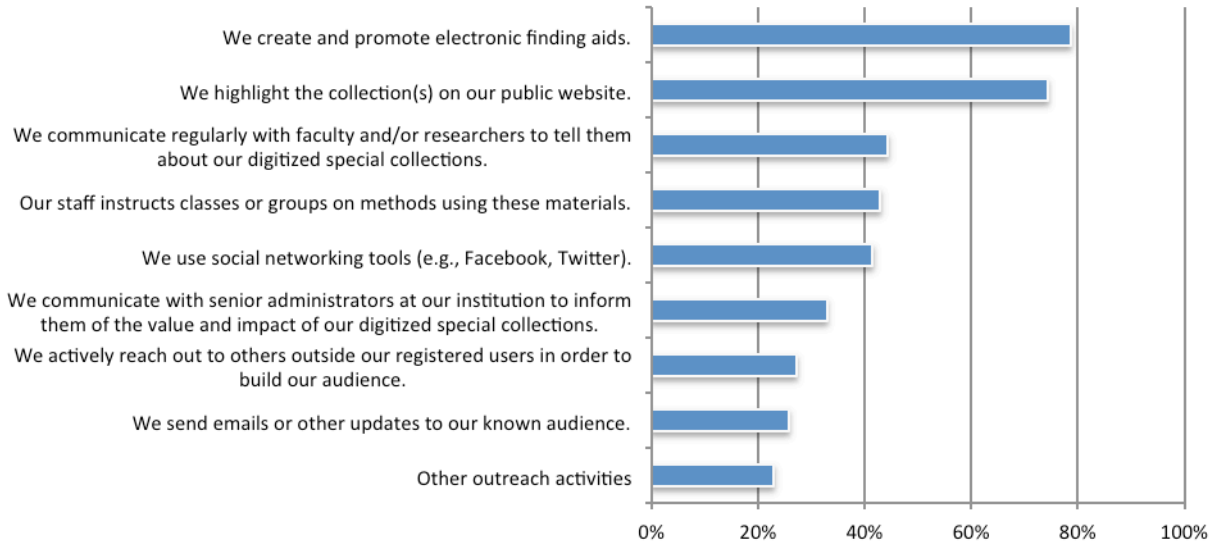
“Building upon the strengths of our work with this collection is where we’re headed, implementing social networking capabilities, where appropriate.”
—An assistant university librarian

provide many alternative information sources for those who seek them, developing robust strategies not just for “passive” discovery but for “active” outreach may be called for. Yet, on the other hand, the outreach efforts of librarians who are aware of the new challenges presented by the digital environment may indeed be constrained by factors such as limited institutional resources or a distributed audience.³⁴

³³ Other research has demonstrated the extent of outreach undertaken by Special Collections librarians. In 2006, ARL conducted a survey of their member libraries in order to gather information about these efforts in Special Collections departments and effectiveness of those efforts. The study was prompted by the belief that “the manner in which Special Collections provide access to their holdings is as important as the substance and extent of the holdings they collect.” For more, see Florence Turcotte and John Nemmers, *ARL SPEC Kit 296: Public Services in Special Collections* (November 2006), <http://www.arl.org/bm~doc/SPEC296web.pdf>.

³⁴ The age of their digitization programs makes no difference in how they seek to engage users, as libraries that have been digitizing longer (beginning in 1999 or earlier) do not perform any specific kinds of outreach with any more regularity than those that are newer to digitization.

Figure 14 Percentage of libraries that *regularly* use the following outreach activities to raise awareness of their digitized special collections (n=70)³⁵



Given the range of activities librarians engage in to ensure that their digitized special collections are discovered and used, little is known about the impact of these activities. Only 17% of respondents reported measuring the effectiveness of the outreach activities they use regularly. It is possible that the other 83% may not simply have the resources to evaluate their impact or they may not know how to do so in the digital environment. As a consequence, these institutions may not know if they are, in fact, reaching and serving the needs of the communities for which they created these collections in the first place.

³⁵ “Other” outreach activities include (in descending order): general PR work/promotional materials, hosting special events, creating exhibits (virtual and physical), publishing on and/or speaking about collections, and reaching out to donors and faculty.

CONCLUSION: STRATEGIES FOR ONGOING SUPPORT

ARL member libraries continue to invest significant financial and staff resources to digitize and support special collections, and most directors believe that these investments will only increase over the next few years. There are many motivations for doing this work, including a practical desire to preserve and protect fragile originals or seeing in these collections a real opportunity to share the riches of an institution with the wider world. The reasons for these investments are important, and it is worth recognizing that the ongoing activities they require are not the same.

Survey data on expenditures and activities shows that, in a given year, ARL institutions are spending significantly more to create new digitized collections than they do for the ongoing support of all of their already-digitized special collections combined. Is this surprising? Maybe not, considering the high cost of scanning various materials, creating metadata, and so forth. But the actual amounts reported were often so low—medians of \$4,550 for usage analysis; \$2,900 for user outreach and support; \$3,209 for editorial—as to suggest that even institutions that have digitized dozens of collections have just a small part of someone’s time devoted to undertaking these activities. Project management—the median for which is under \$14,000 per year—across all an institution’s digitized collections can only mean that most collections do not get much attention at all. At the same time, the presence of a few outliers and those whose respondents reported spending many times as much as everyone else suggest that there may be some fundamentally different strategies at play.

For the most part, ARL libraries are taking what we might characterize as a “maintenance” approach regarding the ongoing support of their digitized collections. In this model, while a collection may be costly to create, it is effectively considered to be “finished” at some point, only requiring basic upkeep to ensure, for example, that the digitized files are present on the website and that links are in working order. Access to the collection is most likely via the library’s website and is sometimes fully integrated into the library’s catalog.

Certainly, some institutions think very hard about this large up-front investment as a key element of their sustainability plan:³⁶ some take pains to carefully curate the collection so that it is comprehensive or develop detailed metadata so that it can easily be searched through a common catalog and identified through network-level search and discovery tactics; others focus on developing platforms that can deliver this content to users and become part of the larger architecture of the library. When asked in this survey to name his most pressing sustainability concern, one director of special collections was unable to identify one: “This project fits well within our digital library infrastructure and does not

³⁶ For examples of institutional strategies for sustainability, see Nancy L. Maron, Jason Yun, Sarah Pickle, *Sustaining Our Digital Future: Institutional Strategies for Digital Content* (Ithaca S+R, 2013), <http://www.sr.ithaka.org/research-publications/sustaining-our-digital-future>.

present any notable sustainability issues to us.”³⁷ The example below offers illustration of this “maintenance” model:

Project A, which lives at a very large, public university, was initiated in 2004 and launched in 2008. It includes 1,500 visual materials and 4,000 pages of text, including manuscripts and printed volumes concerning a specific population. It cost \$160,000 to build, 45% of which was drawn from the library budget; the rest came from donations and philanthropy. Today, the project is managed by someone who also has other responsibilities at the institution. Google Analytics helps them to study users by “location, language, browser type, how a user found the site, how long they stayed on a page” and more. The person reporting on this collection noted that, because of a partnership with a hosting platform, the only activity they themselves need to do to support the collections on an ongoing basis is “troubleshooting” in order to ensure the links to their images are all working properly. For this, the library spends approximately \$250 annually from its base budget.

From this description, one could draw the conclusion that the project is very efficient, requiring just basic maintenance. This is the profile of most of the ARL institutions responding to the survey. And yet a closer look offers a sense of what might be missing:

Project A reports that, although they “regularly” engage in a wide array of outreach activities, it has been difficult for them to learn enough about their users and their needs. “Our audience may be from academic institutions or be independent researchers. They may come from all around the world. We do not know about the research interests of our virtual visitors based on the data we collect, so we do not know if their research interests parallel our collecting intent.” In addition, decisions made early on have made it difficult to develop further. “We would like to add additional content to the site; however, because of the way the site was constructed and is currently hosted, we are unable to do this without significant labor or without redesigning the site’s entire infrastructure.”

A small handful of projects appear to be taking a more purposeful, or “*growth*,” approach, where ongoing investment and development is accepted as critical to maintaining the value of a project over time. This might mean that the project benefits from a designated project manager and that more time is spent in learning about the audience and its needs for and uses of the project, as a way to make choices about which content or technical enhancements may be required to keep the collection useful and valuable.

Project B, which resides at a large, private university, began in 1996 and first went online in 1998. The collection is composed solely of digital facsimiles of manuscripts from libraries and collections around the world. The impetus for creating the collection was very strongly due to user demand and the benefit of “bringing together [a] physically dispersed collection of like

³⁷ The individual projects discussed here are actual collections drawn from Part 3 of the survey, in which respondents were asked to report on a single specific collection. For more of the sustainability concerns shared in Part 3, see Appendix B.

materials.” (These factors were rated 10/10, while preservation was rated 2 out of 10 as a motivating factor.) The cost to create the collection was \$750,000, and it is primarily overseen by an individual who also manages other projects. The collection requires a full suite of ongoing activities. Expenses in the prior year were about \$30,000, half of which went to new scanning, about a quarter to project management, and the balance to fund metadata creation, user outreach and analysis, and preservation, with 64% coming from the base budget and 36% from the project partners. This investment appears to have paid off: the respondent reports that audience for the site has “increased substantially” since last year and believes that this is due to having added much new content during that time. Just 20% of this audience is estimated to come from members of the campus community; another 70% comes from the wider scholarly community worldwide, and about 10% are general enthusiasts.

Going forward, it seems that even greater investments will be needed. The library reports that its greatest challenge is in securing staff (or others) to develop metadata for content on the site. While there seems to be a healthy user community, it is not clear how that community can help to further support the project. Though the library has tried soliciting donations on the website, this has not yet been successful.

Projects A and B typify the low-maintenance approach and the purposeful, “growth” approach, respectively, to managing digitized special collections. Both models have their own appeal: For libraries with the robust infrastructure to ingest, store, and deliver content, new digitization projects could be, in effect, the introduction of new content into a pipeline. By this logic, each individual digitization project must be well executed, but once completed, it becomes a seamless part of an institution’s digital library. And this strategy absorbs equally well the carefully curated “collections” and the more ad-hoc digitization of parts of a larger collection. This is a direction we see many libraries taking, and it is a clever solution to some of the most expensive and complex issues.

Yet in itself, it does not fully resolve the questions of impact, audience, discovery, or, in short, how people will come to learn about the content and come to use it. Where there is a digital library or other shared platform in place, the roles of user interface, outreach, and the development of an understanding of audience falls to whoever manages that larger entity. If there is no digital library, or if the digital library does not take on this role effectively, there may be more investment needed in finding ways to be sure that the digitized special collections are reaching the audience and having the impact that was intended.

Whether at the individual collection or digital library level, keeping pace with the fast-evolving needs, desires, and expectations of the users of their digitized special collections will require further investment as over time, user needs will most certainly change. Off-the-shelf analytics programs may be free, but without the time needed to assess their meaning, they are not terribly valuable. Before a project director can decide whether any enhancements are needed—whether more content, innovative tools, or something else—investment in user research must be made. Some of the more costly methods, including focus groups and interviewing users, can offer ways to more directly understand who is using

the materials and how, permitting a clearer way to make strategic decisions about the sort of enhancements that may be needed.

To this end, we offer for consideration the following recommendations for library directors and those directly involved in the creation and ongoing support of digitized special collections:

- Determine the desired impact of the digitized collections. If digitization is intended to preserve fragile originals, simply having created digital surrogates is likely sufficient. If the goal is to provide content to populate a digital library, there, too, perhaps most work is finished once the digitization phase is complete. But if the goal includes ensuring that the content has some measurable impact for its intended audience, then this will require an additional set of activities to understand user needs, to inform them about this material, and encourage engagement.
- Oversight for digitized special collections is a highly distributed activity, but this does not have to be a liability. It is crucial to clarify the roles of the individuals and departments involved, particularly as a project moves from its initial creation to an ongoing maintenance strategy. For those projects where ongoing attention is important, collaboration between the Special Collections department and the library IT team is critical to integrate the digital curation, preservation, and access activities of Special Collections into the institutions' overall digital infrastructure.
- If the intention is for the collection to have wide and/or deep impact among an audience or audiences, consider investing in qualitative research to learn more from those who are or could be using your collections. At very least, this will provide rich information about the value scholars and students place on the digitized collections. At the very best, it may also offer some useful revelations and suggest future sources of support.
- For those institutions whose digitized special collections demonstrate either high usage or strong, measurable impact among a narrow niche, consider what sources of financial support may be available beyond the core budget of the institution. There may be opportunities, such as advertising or sponsorship, to generate funds in ways that do not compromise the ability to keep content openly available but rather offer some flexibility in budgeting for enhancements.

As libraries continue to seek ways to differentiate themselves and expand the impact of their special collections to people around the world who otherwise may never be able to view the physical materials on site, several viable alternatives have begun to emerge. Whether building a central library and digitizing content to populate it or nurturing the ongoing enhancement of a particularly valuable curated collection, developing a deeper understanding of its users and uses will offers a clearer map to the critical choices ahead.

APPENDIX A—ADDITIONAL NOTES ON METHODOLOGY

The deployment of the survey was a complex process, and the research team took every step to ensure that the conclusions drawn would be as accurate as possible. This section is intended to offer additional detail concerning the survey methodology and data analysis conducted.

The survey instrument comprised three parts. The research team sent the first section to library directors and parts two and three to other people working in the library, based on recommendations from the previous section's respondent. After the survey closed, the data preparation and analysis phase began.

DATA PREPARATION

All data was first downloaded from SurveyMonkey and was organized, coded, and stored in an Excel spreadsheet. Responses from the same institution were matched across the different survey sections to facilitate a fuller analysis.

Almost all of those who began the survey sent to them completed the entire survey; incomplete surveys where respondents only completed basic demographic data were not included. For the small number (3) that only responded to the first few questions of a section, we could not be sure whether they meant for their responses to be understood as complete or whether they intended to return to the survey and perhaps adjust some of their answers; these were not included. Blank answers from respondents who reached the end of their sections of the survey were considered intentional, and included.

Two categories of questions in particular posed some challenges, either due to definitions or the complexity of gathering the requested data. We took the following steps to ensure the best quality of the information we received:

1. Financial data—Many respondents reported that providing cost data was a challenge, since the activities the survey asked about are not often part of a discrete budget. That said, most did attempt to determine what these costs would be, by speaking with colleagues and developing best estimates based on percentages of people's time spent on the activity named. Still, the data gathered included some inconsistent answers, including those who chose to leave a field blank rather than entering a "0." It is impossible to ascertain whether this indicates no activity, that it was too difficult to obtain an estimate, or that the activity is being done elsewhere (by a partner institution, for example). We were able to resolve some discrepancies by asking the respondents for clarifications, but many requests went unanswered. In the end, the data shared in this report reflects the mean, median, high, and low for just those responses where the respondent entered a number higher than zero.

2. “Collections”—The question concerning how many digitized special collections are currently hosted or managed by a respondent’s institution elicited a variety of answers (e.g., “300,000”), suggesting that not every respondent adhered to our definition of a “collection”.³⁸

To verify this data, we either checked online or emailed respondents to confirm or correct some of the more unlikely responses and inconsistent responses (e.g., a respondent who said they had no digitized audio files in their collections who also noted they had completed 100% of the metadata for their digitized audio files).

All reports of Canadian dollars were converted to United States dollars at the exchange rate of the day the data clean up began. Additionally, in order to best facilitate analysis, open-response questions were coded into common categories or were assigned to the nearest relevant answer choice so that they, too, could be analyzed with the other responses. For instance, when respondents reported on the assignment of responsibilities for the ongoing support of digitized special collections, so many of the “other” responses indicated some form of a “digital department” that we created a new category to represent this answer.

Ten-point Likert scale answers were used on questions concerning attitudes and perceptions. Responses were banded for analysis. For example, on a Likert scale where respondents are asked how well a statement describes their institution—where a mark of 1 indicates “Not at All” and 10 indicates “Extremely Well”—findings were grouped in three bands: 1–3 = “not well”; 4–7 = “neither well nor not well,” and 8–10 = “well.”

BASIC DATA ANALYSIS

The first steps of analysis looked at general descriptive data (means, medians, maximums, minimums, and standard deviations and frequencies, where appropriate) for the questions in order to get a general sense of ARL member institutions’ strategies for supporting their digitized special collections.

For some questions, including the reporting on costs and revenue generation, we elected to use medians rather than the mean to describe typical costs, due to the presence of some extreme outliers.

³⁸ The definition of “digitized special collections” provided in the survey: “collections of rare or archival content that would ordinarily be kept in a secure special collections or archives facility that your institution has either digitized itself, chosen to have someone else digitize, or taken on the responsibility for managing (i.e., a collection created elsewhere that your institution now hosts and manages). This includes collections of content that are intentionally created or curated by your staff (e.g., the papers of a famous writer, documents from the civil rights movement). A collection could include multiple format types on a specific topic, or a single content type. Some collections may include some ‘born digital’ elements (e.g., a writer’s papers that include print manuscripts as well as email correspondence). If you consider it a ‘collection’, please report on it here. This definition excludes digitized general collections, digital resources that your institution licenses from outside providers (such as commercial databases), and collections composed entirely of born-digital content, such as general administrative records of your host institution. It also excludes ad hoc patron scanning request for specific documents.”

When examining the mix of funding sources each institution relies upon for creating and maintaining digitized special collections, these are reported as averages since all respondents were confined to a 100% scale when asked what percentage of their funding came from what sources.

With guidance from these findings, we sought out both overlaps and divergences among institutional practices and perspectives in categories ranging from the support from library leaders; size; the age of digitization programs; management style; types of and investment in ongoing activities; funding sources for these activities; and efforts to engage with and understand users. The aim of this method was to identify trends and anomalies that might not show up as significant in statistical analysis (perhaps because of our small sample size), but nonetheless might still be interesting and point to future areas of investigation.

ADDITIONAL STATISTICAL ANALYSIS

Once we began to see trends in the data, we conducted statistical analysis focusing on the perspectives and activities we believe can provide a good environment for the sustainability of digitized special collections. This was done using Stata, and, unless otherwise noted, levels of significance were reported at the 10% confidence level. In accordance with standard practice, the research team used the following tests: t-tests to analyze normally distributed data; Kruskal Wallis tests to analyze independent variables with two or more levels and with an ordinal dependent variable; Wilcoxon-Mann-Whitney tests when comparing the means of two groups with a dependent variable that is neither normally distributed nor interval; one-way ANOVA tests to look at differences between groups of normally distributed interval dependent variables; and a logistic regression analysis when the outcome variable was binary and the predictor variables were either interval or binary.

The points where we noted statistical significance in the main report are discussed below and are grouped by the categories covered in the report.

Representativeness of Sample

Our sample was determined to be representative of the ARL population by comparing the total and material expenditures for the entire ARL population with the expenditures of those who responded to our study. The mean total expenditures of the ARL population was \$35 million, while the mean total expenditures of the responses in our sample was \$31 million. A paired t-test and a two-sample t-test resulted in p-values of .47 and .6; both statistically insignificant. The mean materials expenditures of the ARL population was \$11 million, while the mean materials expenditures of our sample was \$12 million. A paired t-test and a two-sample test resulted in p-values of .91 and .39, respectively, indicating that there is no statistically significant difference between the two groups. Based on these two measures, as well as the fact that all who were contacted were given an equal opportunity to respond, we concluded that we had a representative sample of the ARL population.

Perspective of Library Leaders

There are two significant relationships between the perspectives of library leaders at the institutions and whether their institutions had attempted revenue generation. Using data from question 2 from “Institutional Perspective” and question 26 from “Collections in the Aggregate,” we used a logistic regression to test for significance. More institutions that responded positively to the statement “digitized special collections are critical to our current strategic direction” have tried revenue generation than institutions that responded negatively. Additionally, institutions that agreed with the statement “The primary purpose of digitizing our special collections is to preserve or protect the physical objects” have tried revenue generation less. Wilcoxon-Mann-Whitney tests and t-tests confirmed that these relationships were statistically significant.

There is also one statistically significant relationship between the library leaders’ perspectives on digitized special collections and ongoing activities: whether that institution tracked users. Using a logistic regression, we found that institutions that agree with the statement “our library or institution has invested sufficiently in updates and functionality upgrades for the special collections materials that we have digitized in the past” more commonly track users than their peers do. These results were confirmed by a t-test.

Management Style

There is no statistically significant relationship between the year institutions began digitizing and the way those collections are managed. Using data from question 2 in “Collections in the Aggregate” and question 6a from “Institutional Perspective,” we ran an ANOVA test and t-test to test for differences in the mean year of first digitization for the various management styles. The corresponding insignificant p-values were .89 and .80, respectively.

There is no statistically significant relationship between the number of collections and the style of management. We used data from question 6a from Institutional Perspectives and question 2 from Collections in the Aggregate. A Kruskal Wallis test resulted in a p-value of .72, revealing a statistically insignificant relationship.

Expenditures

A Kruskal Wallis test found that the only statistically significant difference between the amount of money spent on various support activities by older (beginning in or before 1999) and newer (beginning in or after 2000) digitization programs was in up-front spending on preservation. The mean initial expenditures for newer digitizers was \$19,844; for older digitizers it was \$11,265. This test had a p-value of .07.

Funding Sources

There is a statistically significant difference in the sources older (beginning in or before 1999) and newer (beginning in or after 2000) digitization programs draw from to support the initial creation of their

digitized collections. A Wilcoxon-Mann-Whitney test demonstrated that older programs receive less of their funding from the base budget of their institutions than do newer programs; the older crowd gets an average of 63% of the base budget, while newer programs receive an average of 81% (p-value of .009).

Wilcoxon-Mann-Whitney tests also show that newer institutions tend to draw less from donations/philanthropy (a mean of 2%, p-value of .031) and partner organizations (a mean of 1%, p-value .038) than older institutions for the initial creation of their resources. On average, the latter receive 7% from donations/philanthropy and 4% from partner organizations.

There are no significant differences in the sources from which older and newer programs draw to fund the ongoing support of their collections.

Revenue Generation

Institutions that have attempted to generate revenue from their digitized special collections began to digitize an average of two years prior to those that have not. This relationship is statistically significant. We used data from question 1 and 26 from “Collections in the Aggregate” and used a t-test to check for significance. The resulting p-value was .03. The average year of first digitization was 1998 for institutions that had attempted revenue generation, and it was 2000 for institutions that had not attempted revenue generation. Although this relationship is statistically significant, it is small in magnitude.

Institutions that attempted revenue generation tended to have more collections (63 on average) than those that had not (38 on average). A Wilcoxon-Mann-Whitney test revealed a p-value of .03, signifying that this difference is statistically significant. To complete this analysis we used questions 26 and 2 from “Collections in the Aggregate.”

There is no statistically significant relationship between revenue generation and ongoing costs. Because the distribution of ongoing costs is positively skewed, with a number of outliers, we used a Wilcoxon-Mann-Whitney test to look for differences in the medians between the two groups. Using data from questions 26 and 19 from “Collections in the Aggregate,” we performed the test and included only institutions reporting expenditures greater than “0.” The corresponding p-values were .92 and .82 respectively, both of which are statistically insignificant.

Although we have noted several areas for which answers may have been difficult for respondents to provide, the research team believes in the integrity of the data presented here. Respondents were forthcoming and generous with their time, and we have made every effort to ensure that our findings would reflect accurately the information they shared with us.

APPENDIX B—OPEN-TEXT RESPONSES TO PART 3: SPECIFIC COLLECTIONS

Part 3 of the survey asked respondents to share information about just one specific digitized special collection at their institutions, and as a result, the answers provided are not representative of all digitized special collections at ARL member libraries. Nevertheless, answers to two of the more revealing questions are provided below, institution by institution.

Please describe the greatest sustainability challenge concerning this collection, and briefly explain how you have tried to address it.	Please briefly describe any other sustainability challenges, and briefly explain how you have tried to address them.
<p>Obsolescence.</p>	<p>The [REDACTED] Web site and underlying search system should be reengineered every 3-5 years, but has not been touched since 2006. Obsolete image derivatives (MrSid) should be replaced with modern derivatives (JPEG2000). Bibliographic errors should be corrected. New content should be incorporated. New permissions should be obtained. Content should be better correlated to newly published [REDACTED] edition of [REDACTED].</p>
<p>Because the collection is in three different locations online—[REDACTED]—any updates/corrections need to be made in three places. One person manages and tracks these updates.</p>	<p>Migration of content into new delivery platforms and preservation systems. We keep many of our digital collections in the same content management system, so that we can manage the migration and preservation of these collections efficiently.</p>
<p>Because this is a newer collection, we've not experienced any great sustainability challenges yet. That said, the most persistent challenge we face with sustaining digitized collections (and which appeared already began appearing immediately upon launch of this collection) is managing expectations for continued enhancement of the collection. Immediately following launch of the site, project sponsors identified changes and enhancements they desired, which would involve beginning digitization and publication of additional collections as well as making substantive changes to our platform for serving digitized collections. Our approach with this collection was to gather those requests, and table any work until the changes could be scoped, reviewed, and approved as a separate project.</p>	<p>An ongoing challenge with all digitized collections (not unique to this one) is ongoing costs for hosting the collections and storing the archival digital copies. We have begun a pilot of hosting our collections through an outside vendor (to reduce maintenance and labor costs with updating server), but as our collections grow, it is becoming clear that storage costs may need to be considered as a capital investment.</p>

Please describe the greatest sustainability challenge concerning this collection, and briefly explain how you have tried to address it.	Please briefly describe any other sustainability challenges, and briefly explain how you have tried to address them.
The greatest seasonality challenge is maintaining and preserving the video formats. Currently, they are backed up, but there is not active bit level preservation on this format. We're investigating a preservation repository solution for preserving these digital assets.	Another sustainability challenge is maintaining the URLs which the platform for streaming the video file generates since it is a vendor platform.
Preservation of the digital object. Dark storage or master files with metadata and multiple master and access versions stored with metadata.	
Keeping up with metadata for new additions. Each manuscript requires extensive bibliographic descriptions and we don't have staff. We have tried to engage scholarly community with limited success.	Website structure is complex, so platform migration will be difficult. We have good documentation and use open source tools whenever possible.
The greatest challenge is the scanning, which is a result of competing interests from units in the Library. We are working successfully with our digital staff to increase the number of pages scanned by setting up weekly appointments and maintaining a steady, but slow flow of documents for digitization.	We have identified several volunteers who assist with the metadata, which have contributed greatly to the project.
Funding, particularly for IT resources to support digital library and digital preservation services.	
This project fits well within our digital library infrastructure and does not present any notable sustainability issues to us.	
We no longer create silo'ed sites for our virtual projects. Resources created as this project are discoverable in a platform in which most of the institution's digitized special collections materials are indexed and discoverable. This has removed the issue of managing multiple sites, which is very much a sustainability issue.	Programmatic long-term preservation of digitized resources is an ongoing concern, and one that competes with other institutional priorities. Currently, we manage these resources on networked storage and investigation digital preservation strategies.
At this point we're most concerned about digital preservation, and we have not done much apart from talking about it. We recognize that we should be addressing this, but it is difficult to give priority to future needs when budget cuts have limited our ability to meet today's needs.	
Software and access interface maintenance; will need to migrate content to new discovery system when current system becomes obsolete	Working with [departmental] staff (which changes annually) to negotiate when to add content to the site
Scaling up the storage capacity for this and all the other collections we also create and mount.	

Please describe the greatest sustainability challenge concerning this collection, and briefly explain how you have tried to address it.	Please briefly describe any other sustainability challenges, and briefly explain how you have tried to address them.
Digital preservation is the greatest challenge. We have implemented a number of processes and technologies, including digital signatures, persistent citation inks, and the use of extensive preservation metadata based on the PREMIS standard.	The next most significant challenge will be the migration of this content to new formats, specifically jpeg 2000. We do not yet have the infrastructure support to do this migration so, as yet, we haven't explicitly addressed this problem.
The most immediate challenge is obtaining long-term funding to support the expansion of the website and preservation of its content.	
Technologically—ensuring that content in various media (video, audio etc.) continues to run and function in a hybrid web/content management system environment. Migrating to new [REDACTED] iteration; exploring alternatives to Ensemble for video/audio delivery.	Adding new content—we want to build modules around additional topics and are attempting to use graduate student interns to do initial work. Finding ways to get word out to users (in particular teachers and faculty) about content—[REDACTED] opportunities have been fruitful.
Changing technology in Web-based [REDACTED] services and challenges to sustained IT services at the institutional level	
Getting coordinated copies for preservation. We have addressed this with software and through agreements with other organizations.	Eliminating management error and data loss. We are working on ways to avoid and detect data loss.
The lag between the scanning and the mounting on both [website] and [another website]. We are addressing this institutionally, by working towards a fully functioning preservation repository with middleware clients to provide user access. And by integrating more closely [website] with that repository.	
Long-term preservation of the master files. The digital collection is hosted on the [REDACTED], one of the many services offered by this statewide consortium of college and university libraries. While there is a [REDACTED] digital preservation task force, a plan has not yet been recommended, accepted, or implemented, and the current infrastructure is perceived to be inadequate. I am attempting to address the challenge by participating as a member of the task force, and also by being open to the possibility of moving the collection once [library] develops the capacity to provide true long-term preservation.	Delivery in a way that makes clear the collection ownership. Addressed through revision of introductory text. Consideration of additional branding, move from the central [REDACTED] to a separate [university] instance, or removal to [library] [REDACTED] instance.

Please describe the greatest sustainability challenge concerning this collection, and briefly explain how you have tried to address it.	Please briefly describe any other sustainability challenges, and briefly explain how you have tried to address them.
<p>We are involved in a [REDACTED] Network that does not currently have sufficient resources to cache copies of most of our content, including large collections such as this one. To address limitations of the system, we split up the manifests into chunks of less than a terabyte, and developed a [REDACTED] plugin to support harvesting. We are purchasing a second server onto which we can sync content, but at present we do not have any copies of these archival files stored safely in a separate geographical location.</p>	<p>Currently this collection is hosted in [REDACTED], which infers the relationships between files and their collection based on file names. We developed open source software, which enables delivery of digital materials that have been linked into a finding aid, regardless of the delivery method of the finding aid. Also, the links added to the finding aid are persistent URLs, so that if the digitized content must be moved, we need only update the database with the new location. The metadata will not need to be altered.</p>
<p>To make it sustainable it has to grow. Funding to add more content is the greatest challenge—addressed by seeking out grants and advocating for funding from other sources to support the growth of the collection.</p>	
<p>Our greatest concern is for the portions not digitized</p>	
<p>We would like to add additional content to the site; however, because of the way the site was constructed and is currently hosted, we are unable to do this without significant labor or without redesigning the site's entire infrastructure. Currently, another institution hosts the images, while yet another institution manages the metadata.</p>	<p>Users have suggested corrections to metadata within the site; however, we have had difficulty correcting the metadata because another institution manages it.</p>
<p>(updating digital object ID's and finding aid, when collection was reprocessed). NOTE that the enhanced organization, arrangement and preservation of the collection may have come about due to the additional exposure and appreciation of the collection resulting from the [REDACTED] digital archives project. (Additional exposure through digitization resulted in higher demand for the collection materials as well as need for finding aid refinements and enhanced preservation of the materials for safer handling.)</p>	<p>Expectations of digital archives and best practices have evolved since 2006, when the [REDACTED] project went live. The [library] and [REDACTED] has repurposed the digital assets in various ways: 1) as a collection on the [REDACTED] site; and 2) as an online exhibition on the [library] site. Digitization of select items or select pages from volumes is less often considered optimal. The sustainability challenge, and opportunity, lies in repurposing and re-contextualizing digitized resources.</p>

Please describe the greatest sustainability challenge concerning this collection, and briefly explain how you have tried to address it.	Please briefly describe any other sustainability challenges, and briefly explain how you have tried to address them.
<p>The [library] is committed to the endeavor, and provides modest staffing and support. The challenge is not biting off more than one can chew, to work within resources, focus on core resources that merit being put online, focus on core user needs, do not reach out for every opportunity.</p>	<p>There is considerable content and considerable variety made available at [website] yet staffing is modest. To maximize productivity in getting content online, it is important to "Do-It-Yourself" as much as possible, focus on core content that can be digitized within the unit's endeavor and abilities, use micro-staffing via repurposing and plan for future reformatting and migration that will be needed. We will try to pool our digitizing initiatives throughout the different special collections over the next few years.</p>
<p>Our [REDACTED] has a backup but not a digital preservation strategy. I am chairing a Task Force to develop that strategy. We will try to achieve 'bit level' preservation—i.e., 'backup on steroids' with physical copies of this collection existing in multiple places. The collection is backed up in-house on our mirrored storage access network, but re-loading the collection from images and metadata in xml files would be painful and expensive, thus the need for a greater preservation strategy that integrates with the platform [REDACTED] in use.</p>	<p>Feel free to contact me about any questions.</p>
<p>No dedicated budget for digital collections</p>	
<p>Preservation and migration</p>	
<p>We are currently developing a digital preservation repository called [REDACTED]; three copies of masters for this collection will eventually be ingested into [REDACTED]. Currently we have 1 copy stored online and 1 copy on DVD.</p>	
<p>The significant investment in scanning high-resolution TIFF files of each page will only be worthwhile if those TIFFs can be preserved. We have harvested them into the [REDACTED] preservation network, where they are backed up by 6 partner institutions and voted/pollled on by [REDACTED] software.</p>	<p>A recent software upgrade required modification of the collection-level link. The old link has been replaced everywhere we can control, but there are probably dead links out there that we don't know about/aren't able to update.</p>

Please describe the greatest sustainability challenge concerning this collection, and briefly explain how you have tried to address it.	Please briefly describe any other sustainability challenges, and briefly explain how you have tried to address them.
<p>Digital preservation. This project has already generated close to 1 million high-resolution TIFF images. They are sent to us from the vendor initially on hard drives and we have had to develop a workflow by which a) quality control is created, b) access copies generated, c) archival masters are sent to tape and offsite. We are working with our campus [REDACTED] to handle the tape archiving of these files.</p>	<p>Access has been another challenge. Due to a lack of resources, we were not able to keep up with the digitization, and while we have 10,000 books digitized, only 1700 are available online, by "online" we mean with very limited access. We have recently developed a batching mechanism to begin loading books again, but the copyright issues remain problematic.</p>
<p>Ongoing challenge will be to ensure the long term stability and preservation of the digital assets created.</p>	
<p>As this is a collection within a consortial project the only maintenance costs or challenge is funding the annual membership fee.</p>	<p>Adding new content as copyright allows— continue to seek internal grant funds for student help, etc.</p>
<p>Preservation of digital content—are working with campus IT on data storage issues</p>	
<p>Long-term preservation of the collection's digital surrogates. The institution is currently planning to implement a repository infrastructure for all digital objects within the Library. We are awaiting policies and protocols outlining the institution's approach to digital preservation.</p>	
<p>The [REDACTED] collection was the first major acquisition received by the [university], dating to 1918. It came with a sizable monetary donation and was eventually placed in a high-profile building, literally and figuratively. Over the course of time, the physical content received less and less use. Now, use of the digital content has exploded by comparison. Internet Archives use information points to some volumes being accessed thousands of times. [REDACTED] set was so vital that we spun off a separate website just for it. Building upon the strengths of our work with this collection is where we're headed, implementing social networking capabilities, where appropriate, and future technological advancements.</p>	

Please describe the greatest sustainability challenge concerning this collection, and briefly explain how you have tried to address it.	Please briefly describe any other sustainability challenges, and briefly explain how you have tried to address them.
We will need to preserve the digital files and the larger database for years to come. We will need to make sure the storage is a trusted repository and that we can access these files when needed much further down the road.	
Migration and storage of large files and complex objects are the greatest challenges. We are just getting around to implementing a digital preservation strategy.	
We started digitizing newspapers when metadata standards like [REDACTED] did not exist, and we are now faced with retrofitting our collection to follow that standard. Imaging formats present a similar problem. We scanned to TIFF and generated PDF files from them for each article. As a result we have more than 14,000,000 individual files for 1.3 million pages of digitized newspapers. The newer standard is to use one JPEG2000 file for each page with article coordinates built-in, which reduces the number of files stored in a database. We are considering how to convert our collection to this method.	High-resolution digital files of scanned images comprise the largest part of our digital archive—perhaps 40-50 terabytes of 75 terabytes. Storing that much data is a challenge and we are about to contract with University IT to lease cloud storage.
The various formats and content of the collection. Trying to maintain the integrity of the original site while updating the data structure underneath.	Managing a collection that had our digitized content imbedded in a part of it. Largely owned by a faculty member and not the library for over a decade, now that the entire collection was been transferred to the library, we have both our content, other content, and scholarly discourse surrounding the data. It is highly complex and has over 200,000 individual files. Have a look at our documentation referred to in the URL above. It may not be the best fit for your survey. If you would like me to do a different, more straightforward digitization collection I would be happy to do it. The survey seems to be predisposed towards imaging but I included this collection because I consider both imaging and textual markup of sc materials all part of digitization.
None have been encountered so far.	None have been encountered so far.
Storage space for digital assets continues to grow. Will become problematic as we move to HD recording. Metadata creation is labor intensive and lags recording by significant amount of time.	

Please describe the greatest sustainability challenge concerning this collection, and briefly explain how you have tried to address it.	Please briefly describe any other sustainability challenges, and briefly explain how you have tried to address them.
New materials are infrequently added to this collection so getting those into the preservation routines is difficult.	We recently migrated from [REDACTED] to [REDACTED].
Copyright clearance of live festival recordings (multiple performers, some unknown, some with complex or unclear copyright issues). We focused on recordings that we suspected would be low risk, tried to build popular support through strategic cultivation of key performers. Still a struggle because we can provide access to the top artists and most interesting workshops because of the risks of litigation. Costly and time-consuming activity that requires a consistency in staffing but very peripheral to our main mandate as a research facility.	<ol style="list-style-type: none"> 1. Updates to the Omeka software chosen to host the content after grant funding is finished. 2. Establishing a streaming server to provide secure access to the recordings with minimal risk of copying or downloading. An argument for the necessity of a secure streaming server was based on the [REDACTED] case.

APPENDIX C—ONLINE SUPPLEMENT: SURVEY INSTRUMENT AND DATA

The following are links to the three parts of the survey instrument administered by ARL and Ithaka S+R. In addition, please find a link to the complete set of data slides from the survey.

SURVEY INSTRUMENT

Sustaining Digitized Special Collections: Institutional Perspective

<http://www.arl.org/bm~doc/digitized-special-collections-survey-part1-26march12.pdf>

Sustaining Digitized Special Collections: Collections in the Aggregate

<http://www.arl.org/bm~doc/digitized-special-collections-survey-part2-26march12.pdf>

Sustaining Digitized Special Collections: Specific Collections

<http://www.arl.org/bm~doc/digitized-special-collections-survey-part3-26march12.pdf>

DATA SLIDES

Sustaining Digitized Special Collections Report Slides

<http://www.arl.org/bm~doc/digitized-special-collections-report-slides-15feb13.pdf>