

UC Office of the President

CDL and Partner Organizations - Project Publications

Title

Summary of Research: Findings from the Building a National Archival Finding Aid Network Project

Permalink

<https://escholarship.org/uc/item/3tv1m5xj>

Authors

Weber, Chela Scott
Connaway, Lynn Silipigni
Doyle, Brooke
[et al.](#)

Publication Date

2023-05-30

DOI

<https://doi.org/10.25333/7a4c-0r03>

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Summary of Research: Findings from the Building a National Finding Aid Network Project

Chela Scott Weber

Senior Program Officer

Lynn Silipigni Connaway

Executive Director, Research

Brooke Doyle

Senior Project Coordinator

Lesley A. Langa

Associate Research Scientist

Merrilee Proffitt

Senior Manager

Bruce Washburn

Principal Software Engineer

Itza A. Carbajal

Associate Researcher



© 2023 OCLC.

This work is licensed under a Creative Commons Attribution 4.0 International License.

<https://creativecommons.org/licenses/by/4.0/>



May 2023

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 978-1-55653-294-8

DOI: 10.25333/7a4c-0r03

OCLC Control Number: 1376802618

ORCID iDs

Chela Scott Weber  <https://orcid.org/0000-0002-6358-5128>

Lynn Silipigni Connaway  <https://orcid.org/0000-0001-9783-675X>

Brooke Doyle  <https://orcid.org/0000-0002-4318-364X>

Lesley A. Langa  <https://orcid.org/0000-0002-7826-6402>

Merrilee Proffitt  <https://orcid.org/0000-0002-2322-8337>

Bruce Washburn  <http://orcid.org/0000-0003-4396-7345>

Itza A. Carbajal  <https://orcid.org/0000-0001-5316-658X>

This report represents work done on behalf of OCLC Research. When this report was written, Bruce Washburn was Principal Software Engineer at OCLC.

Please direct correspondence to:

OCLC Research
oclcresearch@oclc.org

Suggested citation:

Weber, Chela Scott, Lynn Silipigni Connaway, Brooke Doyle, Lesley A. Langa, Merrilee Proffitt, Bruce Washburn, Itza A. Carbajal. 2023. *Summary of Research: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/7a4c-0r03>.



This project was made possible in part by the Institute of Museum and Library Services, through grant [LG-246349-OLS-20](https://www.imls.gov/grants/2019-2020). The IMLS is the primary source of federal support for the nation's libraries and museums. To learn more, visit www.imls.gov.

The views, findings, conclusions or recommendations expressed in this project do not necessarily represent those of the Institute of Museum and Library Services.

CONTENTS

| | |
|--|----------|
| Executive Summary | v |
| Introduction | 1 |
| Context: How We Developed Our Research Questions | 2 |
| Methodology: How We Addressed Our Research Questions | 4 |
| End user research | 4 |
| Aggregation contributor research: cultural heritage institutions | 5 |
| Aggregation contributor research: descriptive data analysis | 6 |
| Findings: What We Learned from Our Research | 6 |
| End user research | 6 |
| Users of archival aggregation | 7 |
| Why users are doing archival research | 8 |
| How users discover and access material via archival aggregations | 9 |
| Benefits and challenges of using archival aggregation | 11 |
| Aggregation contributor research: cultural heritage institutions | 13 |
| Factors that enable creating archival description | 13 |
| Factors that constrain creating archival description | 13 |
| Incentives to participating in aggregation | 14 |
| Disincentives to participating in aggregation | 15 |
| Aggregation contributor research: descriptive data analysis | 16 |
| Data structure, consistency, and ability to support discovery | 16 |
| Surfacing connections between collections | 17 |
| Access to digital content | 18 |
| Connecting aggregation users to holding archives | 18 |

Recommendations: Applying Research Insights to the NAFAN Project19

- Value proposition for a national archival aggregation 19
 - Address discovery challenges..... 20
 - Address access challenges 20
 - Address archivists’ challenges 20
 - Address inequity in the archival landscape 21
- Potential challenges to sustainability22

Conclusion 23

Acknowledgments 24

Notes 26

EXECUTIVE SUMMARY

Researchers face significant challenges to finding and using archival collections relevant to their interests. Many archives struggle to publish their finding aids online. Those archives that do have information about their holdings online often only make these available on their website, requiring researchers to conduct repetitive inquiries in siloed systems. Regionally based aggregators of archival description have contributed significantly to making archives more visible on the web. Still, only 14 aggregators covering archives in 23 states currently exist in the US.

In 2020, IMLS awarded the California Digital Library (CDL) a National Leadership Grant to support Building a National Finding Aid Network, a project to build the foundation for a national archival finding aid network to address issues in the current archival discovery landscape (LG-246349-OLS-20). CDL led the project with partners at OCLC, the University of Virginia Library, Shift Collective, and Chain Bridge Group.

OCLC was responsible for conducting research to illuminate the needs of both archival researchers and cultural heritage institutions regarding finding aid aggregation and evaluating the quality of existing finding aid data at scale.

Research Overview

OCLC developed research questions scoped to address project priorities. Data gathering and analysis included:

- A survey placed on the sites of 12 regional archival aggregators, yielding a total of 3,352 usable responses from end users
- Semi-structured individual interviews with 25 end users
- Focus group interviews with 53 archivists
- A quantitative analysis of over 145,000 EAD encoded finding aids, collectively representing 741 repositories

This report contextualizes and synthesizes the findings across OCLC's research activities, focusing on how findings relate to future work on the NAFAN project. Findings detail:

- **End users:** who the current users of archival aggregations are, why and how they are trying to discover and access archival collections, and the benefits and challenges they face when searching archival aggregations
- **Cultural heritage institutions:** what makes it easier or harder for organizations to describe the archival collections in their care, what value participation in archival aggregation might bring to organizations, and what may influence whether organizations contribute to an aggregation
- **Descriptive data analysis:** how much consistency exists across the body of EAD encoded records found in current aggregations of archival description, and how current data would support user discovery and access needs

Recommendations

OCLC's research findings highlight a variety of concerns and potential directions for a future National Finding Aid Network and suggest functional and programmatic considerations for future phases of work.

VALUE PROPOSITION FOR A NATIONAL ARCHIVAL AGGREGATION

A core principle guiding work on this phase of NAFAN is that the project must gather evidence to better understand and articulate the value proposition that a national aggregation can provide.

Alignment between the needs of end users and archives illuminates how a national aggregation can provide value by addressing the urgent shared needs around the visibility and discoverability of archival collections. These include addressing researchers' difficulty in discovering archival collections on the web and accessing the archival materials they identify and archives' challenges in creating and publishing archival description and making their archival collections visible on the web.

Our research also identifies a significant disparity in the resources available to both archives and archival researchers, which impacts the historical record, scholarship, and knowledge access and dissemination. Addressing the broad needs of disparate archives and archival researchers would be an important intervention in an uneven access landscape.

POTENTIAL CHALLENGES TO SUSTAINABILITY

OCLC's research points to significant misalignment between resources and expectations that will challenge sustainability.

First, there is a tension between attitudes around aggregation data requirements, users' discovery desires, and archivists' discovery expectations. Archivists identified their top barrier to contributing to aggregations as having to comply with the aggregator's data requirements. At the same time, archivists and end users expressed discovery desires and expectations that require structured data. Discovery benefits were archivists' most-cited incentives to participate in aggregation, but archival description resources and practices do not align easily with the discovery functionality both users and archivists described as desirable.

Second, but no less important, is that archivist expectations for low- or no-cost participation do not realistically align with their desires for robust system functionality. The features and community support desired by archivists would require significant monetary support, and contributors would likely need to bear the majority of ongoing costs. Further investigation is required to assess what costs the archival community is willing to bear and how market segmentation could shape a fee structure that is equitable across many types and sizes of institutions.

Conclusion

Our findings indicate that there is significant value to be drawn from a national aggregation of archival description; however, there are significant challenges to overcome to build the community of participation required to support sustainability. This research presents concrete guidance for subsequent phases of the NAFAN project for technical requirements, design and usability, governance, marketing, community building, and further research.

INTRODUCTION

Researchers of all types face significant challenges to finding and using archival collections relevant to their interests. Many archives struggle to publish their finding aids online. Those archives that do have information about their holdings online often only make these available on their local website, requiring researchers to search across many siloed catalogs and websites to do their research. Regionally based aggregators of archival description have made a significant contribution in making archival description more visible on the web. Still, only 14 aggregators covering archives in 23 states currently exist in the US.¹

In 2020, IMLS awarded the California Digital Library (CDL) a National Leadership Grant to support Building a National Finding Aid Network, a two-year research and demonstration project to build the foundation for a national archival finding aid network to address the inconsistency and inequity of the current archival discovery landscape (LG-246349-OLS-20). CDL led the project with partners at OCLC, the University of Virginia Library, Shift Collective, and Chain Bridge Group.

OCLC was responsible for conducting and synthesizing research to understand the needs of both archival researchers and cultural heritage institutions regarding finding aid aggregation. OCLC also undertook an evaluation of the quality of existing finding aid data at scale to scope the network's initial functionality and lay the groundwork for iterative data remediation and expanded network features in subsequent project phases.

OCLC's research findings indicate that there is significant value to be drawn from a national aggregation of archival description. There is much alignment between the needs of end users and archives, pointing to clear ways that a national aggregation can address the urgent needs of both audiences related to the visibility and discoverability of archival collections. A national aggregation of archival description could also make an important intervention in the current inequitable landscape of access to archival materials.

OCLC's research findings indicate that there is significant value to be drawn from a national aggregation of archival description.

The research findings also point to significant challenges that must be overcome to build the community of participation that a national finding aid aggregation will require to be sustainable. Considerable heterogeneity in archival metadata will create data remediation challenges and impact discovery functionality. Archivist expectations around cost and ease of participation do not realistically align with their desires for robust system functionality.

This summary report contextualizes and synthesizes the findings of all OCLC's research activities for the grant, with a focus on how findings relate to future phases of work on the NAFAN project. The research described in this report provides valuable insight that can be used by the NAFAN project and inform other projects serving archival researchers and archivists, those wishing to leverage archival descriptive data, and researchers of archives and archival users for exploring new paths of inquiry for their own research. Detailed reports of findings are available for each individual data collection and analysis activity; readers are encouraged to examine these for significant additional information on methods and findings.²

Context: How We Developed Our Research Questions

To shape research questions that best addressed NAFAN project goals and contributed to the broader professional knowledge, the research team reviewed project documentation and performed a review of relevant archival studies literature. The result was a series of research questions scoped to address clear needs for the project and important gaps in the archival literature.

A primary driver for the project's research was "Toward a National Archival Finding Aid Network," a 2018–2019 planning initiative led by the CDL and supported by IMLS through Library Services and Technology Act funds. The initiative convened a symposium of US finding aid aggregators and expert advisors to explore the potential for a national model for archival aggregation. Symposium participants called for a substantially more inclusive framework for aggregation that meets the needs of a more diverse set of end users, enables participation by a broader range of cultural heritage institutions, and supports a greater variety of collection description levels and formats while simultaneously transitioning away from outmoded technologies and directly addressing foundational issues of sustainability.³

Outputs from the symposium included a summary report of findings and an action plan outlining a concrete path toward a set of goals.⁴ OCLC examined these outputs to identify the research questions they raised; OCLC and CDL teams then prioritized and narrowed the questions for inclusion in this study. After questions were prioritized, the OCLC research team conducted a literature review to identify any existing work that addressed or would inform the research questions.

The literature review identified a significant gap in existing archival research, with no studies examining researcher needs related to aggregations of archival description that represent multiple institutions' holdings. Much early work focused on researcher interaction with and understanding of finding aids as a genre, or with finding aids as rendered in online interfaces. This work has largely focused on the use of and interaction with the collections and interfaces of a specific institution.⁵

More recent archival user study work has broadened in scope to determine the information-seeking behaviors and research practices of scholars using primary sources, but it has focused almost entirely on historians and other academic researchers in the humanities.⁶

Recent work on archival user personas in service of archival software development projects indicates that users of archives are not only scholars, but also genealogists, local historians, and avocational researchers; K-12 educators; a range of researchers using archives for their professional and creative work in fields such as journalism, documentary filmmaking, and fiction writing; and public services librarians and archivists.⁷ There is little literature on this broader cross-section of archives users. Existing work indicates different researcher types have significantly different needs.⁸

Given the lack of understanding of archival aggregation users, and of the broad range of archival researchers generally, exploratory work was necessary to identify the needs, expectations, and behaviors of potential national aggregation users. The following research questions were developed to address this need.

End user research questions

- Who are the current users of aggregated archival description? Do current user types align with the persona types and needs identified in recent archival persona work?
- Why are current users trying to discover and access archival collections via aggregation of archival description?
- How are current users discovering and accessing aggregations of archival description? What are the benefits and challenges users face when searching archival description in aggregation?

There is no extant research specifically focused on contributing finding aids to an archival aggregator. More broadly, there is a body of research related to the use and implementation of the EAD encoding standard. This work is relevant to NAFAN as EAD encoded finding aids comprise most data in current regional aggregators in the US⁹ and these regional aggregators are likely to play a key role in contributing data to a national aggregation platform. Most of the EAD-focused literature dates from the late 1990s and early 2000s in the nascent period of implementation of the EAD standard.¹⁰ Two studies from 2013 analyzed bodies of EAD encoded findings aids to understand the completeness and consistency of EAD tag usage and how EAD can support online discovery.¹¹

With no existing research about working with archival aggregators, it also was vital to identify and document the experiences of current and potential contributors to archival aggregations.

Given the age of these studies, and the subsequent significant development and uptake of archival collection management systems such as ArchivesSpace and Access to Memory (AtOM), the research team felt it was necessary to develop a current understanding of EAD use through

both an organizational and technical lens. Additionally, given the goals outlined in the 2018–2019 planning grant to enable participation by a wider range of cultural heritage institutions and support greater variety of collection description levels and formats, it was important to develop a broader understanding of descriptive practice beyond EAD creation. With no existing research about working with archival aggregators, it also was vital to identify and document the experiences of current and potential contributors to archival aggregations.

The research questions identified to address these issues were divided into two subsets, one that focused on the cultural heritage organizations that would contribute archival description to a national aggregation and the archivists who work in these settings, and the other focused on archival description as data.

Aggregation contributor research questions: Cultural heritage organizations

- What are the enabling and constraining factors that influence whether organizations describe the archival collections in their care?
- What are the enabling or constraining factors that influence whether organizations contribute to an aggregation of archival description?
- What value does participation in an archival aggregation service bring to organizations?

Aggregation contributor research questions: Descriptive data analysis

- What is the structure and extent of consistency across the body of metadata records in current aggregations of archival description?
- Can that body of metadata records support user needs identified in findings from the user research phase of the study? If so, how? If not, what are the gaps?

Methodology: How We Addressed Our Research Questions

Given the large number, broad scope, and exploratory nature of these research questions, OCLC chose a mixed methods approach to this study.¹² The research team gathered and analyzed both quantitative and qualitative data and considered the data sets in relation to each other. All research with human subjects adhered to the requirements of the UCLA Institutional Review Board (IRB). All survey and interview protocols received clearance from the UCLA IRB prior to data collection.

Further detail follows on the methodology employed in each of the major areas of focus: end user research, aggregation contributor research with archivists working in cultural heritage institutions, and aggregator contributor research with descriptive data.

End user research

For the end user-focused research questions, OCLC employed a mixed methods design, first collecting and analyzing quantitative data via a pop-up survey and then using these data to select participants for semi-structured individual interviews. This method was chosen because very little

was known from existing research about users of archival aggregation. Employing a survey first helped us learn broadly who was using archival aggregation and then make informed selection about who to include in individual semi-structured interviews. Semi-structured interviews were used to collect more detailed information about how and why individuals use finding aid aggregations.

A pop-up survey was used to gather information directly from users and was placed on the sites of 12 archival aggregators who served as partners on the NAFAN project. The survey appeared on the aggregator portal's homepage, on search results pages, and on the landing page for each finding aid published within the aggregator site. This approach captured users no matter how they came to the aggregator website, whether visiting the site directly or following a link from another website or search engine results. The data collection ran from March to May 2021. Respondents were given the option to enter a raffle for a \$100 gift card incentive.

The total survey response was 3,352 usable responses across all aggregators. The pop-up survey data were analyzed using descriptive statistics and a comparative analysis of the responses from the different types of users identified in the demographic data.

A Two-Step cluster analysis was performed on the pop-up survey data to identify groupings of users with strong similarities. The five clusters of users resulting from the analysis were:

- Archivists, librarians, and other professionals
- Faculty and others with doctoral degrees
- Undergraduate and graduate students
- Family history researchers
- Personal interest researchers

Participants from each cluster were then invited to participate in the semi-structured individual interviews. A total of 25 interviews were conducted and recorded, with five users from each cluster participating in the interviews. The interviews occurred between October and December of 2021. Participants were offered a \$50 gift card as incentive for their voluntary participation in the 45 to 60-minute interviews. Using NVivo software, transcripts of the recorded interviews were analyzed using inductive and deductive coding and a content analysis approach.

Aggregation contributor research: cultural heritage institutions

Focus group interviews were chosen as the data collection method to address the project's research questions related to cultural heritage organizations that might contribute to an archival aggregation. Although focus group interview data cannot be generalized to an entire population, the methodology is frequently used for identifying perceptions and attitudes of a target population,¹³ which was the purpose of the focus group interviews in this study.

Focus group interviews were conducted with 53 archivists. Participants came from both institutions that currently participate in an archival aggregation and those that do not, and represented a variety of job responsibilities, sizes and types of institutions, and years in the profession. Focus group interviews ran for 60–90 minutes and were recorded. Participants were asked questions

related to creating and publishing archival description, and to participating in archival aggregation. Content analysis was used to identify the emerging themes from the transcripts of the focus group interviews and the responses were coded using themes.

Aggregation contributor research: descriptive data analysis

Archival data analysis used a mixed methods approach, primarily collecting and doing quantitative analysis on a corpus of EAD encoded archival data and informing that analysis with quantitative and qualitative data from other project research activities. Two phases of quantitative analysis were conducted on the same corpus of EAD data. The first phase used questions determined at the outset of the research, focusing on the structure and consistency of the data and how it could support basic discovery functions. The second phase of analysis used questions focused on how well the data could support specific discovery, access, and back-end system needs, which were surfaced by both the end user and cultural heritage organization research. This approach allowed the research team to take advantage of the other data collection and analysis efforts occurring across the project, and to address emerging themes or needs surfaced within that research.

The quantitative analysis was conducted on a corpus of EAD encoded collection descriptions provided by current finding aid aggregators. Twelve regional aggregators of EAD finding aids participated in the NAFAN project and made their finding aids available for analysis. The resulting data set is composed of 145,673 EAD XML files, collectively representing 741 repositories. The data were analyzed using a variety of computational tools and methods including XPath language queries for selecting nodes from XML documents; R software for statistical computing and graphics; OpenRefine for cleaning, analyzing, and reconciling data; and Microsoft Excel for summarizing data and generating visualizations.

Findings: What We Learned from Our Research

The following section provides a high-level summary of findings across OCLC's data collection and analysis efforts and focuses on how these findings address project research questions. More detailed reporting on findings is available for each individual data collection and analysis effort.¹⁴

End user research

The research questions guiding OCLC's work with end users focused on who current users of aggregated archival description are and whether they align with recent archival persona work, why and how current users are trying to discover and access archival collections via aggregation of archival description, and what the benefits and challenges are that users encounter when searching archival description in aggregation. A discussion of the findings related to these questions follows, drawn from both the pop-up survey and the semi-structured individual interviews.

USERS OF ARCHIVAL AGGREGATION

The pop-up survey provided broad demographic data about the group of respondents using archival aggregators. The age distribution of survey respondents ranged from 18–65+ years old. The highest percentage of the survey respondents were 65 and older (34.4%), with representation falling slightly with each age group. Users of archival aggregation were largely well educated, with 85.5% of pop-up survey respondents indicating they had attained some level of higher education. The largest proportion of respondents held a master's degree (35.6%), 27.1% had a bachelor's degree, 15.5% a doctorate degree, and 7.3% had attained an associate's degree.

Users of archival aggregation were largely well educated, with 85.5% of pop-up survey respondents indicating they had attained some level of higher education.

When asked about profession, the highest portion of survey respondents (21.4%) reported that they had retired from full-time employment. The next highest ranked profession is information professionals; librarians and archivists made up 13.9% of respondents. Graduate students and faculty each made up a little less than 10% of respondents (9.9% and 9.6% respectively), and genealogists represented 8.6% of respondents. The professions representing less than 7% of respondents included undergraduate students, journalists and writers, artists and filmmakers, lifelong learners, museum professionals, K-12 educators, historians, and independent researchers.

A cluster analysis of the pop-up survey data identified five groupings of users with strong similarities—archivists, librarians, and other professionals; faculty and others with doctoral degrees; undergraduate and graduate students; family history researchers; and personal interest researchers. These clusters are aligned in many ways with the persona work done in service of recent archival software projects.¹⁵ They do not introduce any new user types not already represented in some way across these personae.

The cluster analysis does, however, suggest some departure from past persona work. In many of these projects, personal or avocational research and genealogy research are combined into one researcher type.¹⁶ OCLC's analysis indicates that genealogists and general personal interest researchers may have enough distinguishing differences to warrant thinking about them as distinct user types. Conversely, archivists have been treated as their own persona across these projects, in some cases with granular personas created for archivists with different roles and responsibilities.¹⁷ OCLC's analysis identifies that the needs of archivists may more closely align with other professionals than previous work has highlighted. Archivists and librarians account for 51.0% of the archivists and other professionals cluster while journalists/writers (7.2%) and professionals (i.e., lawyers) (10.6%) represent 17.8% of the cluster.

WHY USERS ARE DOING ARCHIVAL RESEARCH

Data from both the pop-up survey and individual interviews indicate that users are seeking archival materials for a range of academic, personal, and professional uses. In the pop-up survey, when asked about the purpose of their research, respondents could pick multiple answers. They ranked personal interest highest at 32.7%, followed by family history research at 24.4%. Professional projects (22.8%), long-term projects (20%), and local history (19.3%) round out the top five selections. All 25 interview participants indicated searching for archival materials for either work, academic, or avocational activities. Many participants pursuing work or academic activities explained that specific final products such as news articles, books, documentary films, class assignments, or dissertations influenced their search for archival materials. Participants seeking archival materials for use in avocational activities were not motivated as frequently by final products, often describing their activities as driven by personal interest, passion, curiosity, or personal growth.

Data from both the pop-up survey and individual interviews indicate that users are seeking archival materials for a range of academic, personal, and professional uses.

In individual interviews, the 25 participants shared a variety of information needs and other motivators for their archival research. Almost all participants (23) described information gaps as motivating their research, which were defined as a need to fill a knowledge gap or identify new information related to a topic of interest. Less than half of participants (10) were motivated by a need to confirm or validate information they already had.

Most participants (21) described a need for contextual information. For some, this meant general historical context to help them better understand the larger picture about a time or place related to their area of focus. For others, this meant finding details or anecdotes to round out a story or help them understand what life was like for the subject of their research. Many participants (19) also described an exploratory information need, often early in the research process. For some, this meant developing initial understandings when working on a new topic or trying to find out if a particular person or organization has records in an archive. For others, it was less directed, and involved doing broad searches to see what information might turn up and how they might use it. Many participants (17) also described a need for factual information such as birth, death, or marriage dates, dates or locations where events took place, correct spellings for names, or full meaning of acronyms. Some participants (14) were looking for evidence to analyze and interpret to build a case or a story, identify patterns or connections, or otherwise build upon to make arguments supported by primary sources. This was most often invoked as a need by those creating new work to be shared with others such as books, articles, plays, documentaries, or museum exhibitions.

All participants (25) described emotional motivations for or reactions to their research, though what provoked their emotion varied. For many, the process of researching brought joy, excitement, or satisfaction. Others enjoyed the thrill of looking for and successfully finding useful material. For other participants, emotional reactions stemmed from interacting with the archival materials itself,

often because the content was personally meaningful or physically interacting with the material was moving. Some participants were motivated to find archival materials that would provoke an emotional reaction in others, help to tell a story, or make historical events come alive for an audience. Relatedly, sharing access to primary sources or the knowledge gained through primary source research was described as a motivator by those pursuing research for both work and avocational activities (15). About half of participants (12) described being motivated by obtaining a reproduction of archival materials.

In terms of what kinds of material might address these needs, more than half of pop-up survey respondents indicated that they are interested in any type of material relevant to their topic (55.8%). Analysis for material preference by subgroup shows that no single group is driving this answer. Retirees, faculty and academic researchers, graduate and postgraduate students, and archivists and librarians note that any material is their preference at 50% and higher. Genealogists want personal family papers slightly more (67.2%) than any material related to their topic (60.2%).

In terms of what kinds of material might address these needs, more than half of pop-up survey respondents indicated that they are interested in any type of material relevant to their topic (55.8%).

HOW USERS DISCOVER AND ACCESS MATERIAL VIA ARCHIVAL AGGREGATIONS

Users are finding aggregators in a variety of ways. Only 20.3% of survey respondents stated that they came to the site because they had used it before. Almost half (44%) found the aggregator through a search engine, and 20.6% reported they found the archival aggregator by following a link on a website or social media. In terms of how often respondents are using the aggregator, 55.4% indicated that when they completed the survey, it was their first visit to that aggregation website. A very small percentage (2.8%) reported using the archival aggregator website daily with more frequent users reporting using the website less than monthly (16.3%) or monthly (12.1%). Interview participants were especially likely to use a regional archival aggregator when their research topic was related to the state or geographic region in which the aggregator was located.

While not specific solely to aggregators, participants praised the usefulness of rich and robust description of archival materials and collections as a crucial tool for archival discovery. Participants identified multiple aspects of the collection description that they found useful. For some, high-level information about an archival collection was important. Information about the size and scope of the collection helped them assess how much might be relevant to their research; information about access restrictions or materials held in off-site storage helped them assess if they would be able to use the collection and what next steps to take to access it. For others, description of collection contents was especially useful, both detailed information such as names of correspondents or granular date information, as well as more robust collection or series-level information that helped them understand the collection as a whole.

Participants also identified repository-level description such as the scope or history of collecting as useful to discovery by helping them identify repositories that might hold unprocessed or uncataloged collections that would be relevant to their research.

Survey respondents are using a variety of discovery tools in addition to archival aggregators, including search engines (62.2%), archive websites (54.4%), genealogy websites (29.4%), university library websites (24.4%), Wikipedia (21.1%), public library websites (16.5%), and WorldCat (13.6%) in their research journey. Interview participants echoed and expanded these findings. When discussing their research process, participants described using multiple social and technical tools to find archival sources, including archival repositories and archivists; their own networks of colleagues, friends, family, or acquaintances; databases supplying access to curated archival materials; citations from secondary sources; and library catalogs.

Once archival material is identified, it must be accessed. When asked about their preference for accessing archival material online versus in person, nearly half of the pop-up survey respondents (42.7%) indicated that they preferred online materials but were willing to use in-person materials. Roughly a quarter of the respondents (23.6%) indicated they had no preference between online or in-person materials. Fourteen percent of respondents stated a strong preference for online only (14.4%) or prefer in person to online (14.7%).

When asked about their preference for accessing archival material online versus in person, nearly half of the pop-up survey respondents (42.7%) indicated that they preferred online materials.

Interview participants described using three different means of access—online, mediated, and in person. Online access to digitized material was the most noted feature helping participants see and use archival collections. Participants cited online access as helping to alleviate barriers because of travel, expense, or time required to do in-person research. Some participants only pursue materials they can use online, while some use online access to documents in preparation for in-person research. Mediated access, or the ability to request either research or reproduction services was another important access enabler, especially for material that had not already been digitized or that was located outside the researcher's local area. Participants who had accessed archives in person described doing so because it was the only way to see the materials they needed, and/or because of a preference for in-person research.

In discussing access to archival collections, many participants described resource limitations as impacting their ability to do research. Cost is a key consideration when deciding to do in-person research in an archive when travel is required. Some participants doing work or academic-related research were able to rely on institutional support for research travel or could extend conference or other work-related travel to include a few research days in an archive. Avocational and many academic researchers had to rely on personal resources to fund research travel. Arranging

for relief from caregiving responsibilities also factored into cost equations and navigating these responsibilities in general impacted research decisions. Cost, along with the amount of bureaucracy involved in making a request, also plays an important role in deciding whether to request reproductions or license archival material for use. Participants usually described assessing a combination of factors when making access decisions, typically weighing available resources against the potential importance of particular archival material to their research.

These barriers to in-person access, along with the limited amount of material accessible online and uneven services for users provided by archives, led to inconsistency in who accessed archives and inequity in how easily they were able to do so. Participants identified their networks and relationships as enabling multiple types of access to collections. This was most often cited by participants in academic or professional settings aligned with archival repositories. In some cases, networks or relationships allowed users to skirt restrictions or delays. Some participants described using peer and colleague networks to get access to content in subscription databases or calling in favors to avoid the long turnaround time of an archives' reproduction or research queue.

BENEFITS AND CHALLENGES OF USING ARCHIVAL AGGREGATION

Though all interview participants were originally identified via a pop-up survey on a regional archival aggregator website, not all of them use aggregators regularly or even remembered using the site where they had filled out the survey. Participants who did use aggregators with some regularity described encountering varying benefits and challenges using the systems. These are discussed here, along with some general benefits and challenges of searching for archival collections that have relevance to the NAFAN project.

The primary benefit of searching for archival collections described across interviews was finding archival material useful to their research. This included finding known collections or items, as well as the serendipity of finding material they did not know existed. Participants discussed the benefit of the wide variety of types of institutions aggregators include, especially when it provided them awareness of collections or institutions they might not otherwise know about.

A key challenge voiced by interviewees is how difficult it is to do archival research generally, including how many places they must look and how long searching takes. The convenience of searching across many archival collections and institutions was commonly cited as an important motivator and benefit of using aggregators because it saved the researcher time and effort. Frustration with having to use multiple systems also was an issue in using aggregators, as regional aggregators have limited scope and multiple systems must be used to approximate a nation-wide search. The most frequently expressed desire among individual interviewees (18) for a future aggregation system was one that had comprehensive coverage of multiple types and sizes of archival institutions across the United States.

A key challenge voiced by interviewees is how difficult it is to do archival research generally, including how many places they must look and how long searching takes.

Relatedly, participants described their desire to better understand what was included in aggregation systems. Researchers were unsure of overlap or coverage across different systems. In many cases, this led them to double-check their searches in other tools or on the websites of individual archival institutions, diluting the benefits of aggregation. Some participants understood that unprocessed collections or those without electronic finding aids would not be included in the aggregation and wanted to better understand how this might impact their search results. Some participants voiced concern that they were not finding material held at smaller and community-based archives and missing out on the stories that those repositories documented. Participants felt a nationally scoped system would save time and effort and help them to feel confident in locating a broad representation of materials.

Participants repeatedly described the aggregators' search capabilities as "powerful." In some cases, this was because the breadth of coverage alongside the ability to sort and narrow results helped them to easily identify resources. In other cases, they discussed the aggregators' capacity to make connections or surface relationships across collections and repositories. Some participants (8) expressed a desire for features in a future aggregation system that would support discovery of connections between collections, such as a recommender system or a network graph type of interface showing related collections.

Interview participants also identified system-related issues as general discovery challenges for users. Confusing interfaces were high on this list, which made it difficult for users to navigate to what they needed, understand what they were seeing, or ascertain relationships between results. Many participants described parsing result sets challenging when searching for archival collections, either because of the volume of results or not understanding their relevance. Participants wanted a way to limit or narrow result sets, and an interface that allows them to quickly view and assess which results are useful to them. A desire for functionality that would help researchers more accurately search was mentioned by multiple participants (10). This included advanced search options to use when constructing a query, as well as filtering or faceting functions to allow them to narrow result sets. Desired fields included geography, type of material, date or date range, keyword, subject, and author or creator.

Another aggregator-specific challenge described by multiple participants was figuring out how to access a collection or take a next step once they identified something of interest. This was especially challenging when the transition from the aggregator to the website of an archival repository did not behave as they expected or wanted, or when the researcher would have to repeat their search at the archival website to find the material they had just identified through the aggregator. Relatedly, many researchers were aware that archives often hold materials not visible online, so they want to be able to easily contact an archive to ask questions about their research. Researchers wanted easy access to contact information, reproduction and reading room photography policies, and information about visiting or requesting research appointments. Participants described aggregators being most beneficial to them when they were intuitive or easy to use and offered easily accessible policies and contact information.

Because of the difficulty of accessing archives in person, many participants (17) cited a lack of availability of digital collections as a general access challenge with real impact on their research. Increased access to digital materials was another major desire mentioned by participants (17) when describing what an ideal national aggregation would provide. For many participants, being able to filter out results without digital access was important.

Aggregation contributor research: cultural heritage institutions

The research questions guiding OCLC's work with institutions who might participate in a national archival aggregation focused on the enabling and constraining factors that influence whether organizations describe their archival collections, the value that participation in an archival aggregation might bring to organizations, and what might influence whether organizations contribute to an aggregation of archival description. The findings reported here are drawn from the focus group interviews with archivists.

FACTORS THAT ENABLE CREATING ARCHIVAL DESCRIPTION

The most frequently mentioned enabling factor that supports archivists in creating archival description was the use of an archival collection management system (CMS). A primary benefit of using a CMS is that it can significantly ease the complexity of creating and publishing EAD; this complexity was noted as a significant barrier to creating and maintaining archival description. Another beneficial aspect of using a CMS is the inclusion of tools that support import and conversion of legacy description.

Another enabling factor for archivists is the use of worksheets or templates to assist the process of creating archival description. While not fully developed CMS solutions, these tools can help staff create descriptions that comply with DACS or ISAD(G) or can be an aid to creating EAD encoded finding aids. Similarly, other types of documentation such as manuals, guidelines, and procedures for creating archival description or using specific tools were mentioned as enabling factors. These tools often were borrowed from or based upon those at another institution and used to support both training and ongoing work.

FACTORS THAT CONSTRAIN CREATING ARCHIVAL DESCRIPTION

The primary challenge that archivists face in creating description is contending with workflows that can be characterized as complex. Participants frequently referenced utilizing multiple systems, managing description in multiple file formats, taking multiple steps to produce descriptive outputs, and creating bespoke workarounds for the limitations of the systems in which they are creating, encoding, and publishing archival description. Many participants specifically described complex workflows to address the need to create both archival and bibliographic forms of description for archival collections, or to manage legacy description¹⁸ across a range of sources of information and in differing file formats.

Participants at institutions that do not have a CMS described multistep, multitool, manual workflows designed to produce EAD encoded finding aids. Participants that did have a CMS with which to produce EAD still described multistep workflows necessary to transform the EAD produced by that system to meet the requirements of a separate publication or discovery system. Some focus group interview participants specifically identified creating or maintaining EAD as challenging. In some cases, this meant that institutions are not creating EAD finding aids because it is too difficult.

Adding to this complexity of systems and workflows is the "never done" nature of archival description, with many participants describing their need to revise descriptions on an ongoing basis. Reasons for this include both revisiting to include new information, such as expanding minimal collection-level descriptions or reparative description projects to remediate harmful language, and revising older descriptions to align with modern-day standards, such as bringing descriptions into compliance with DACS or migrating formats.

Resourcing factors also create challenges to doing descriptive work in archives. Archivists are challenged to find both staffing and budgetary resources to move descriptive work forward, especially when balanced against other competing priorities at their institution. Some archivists specifically brought up insufficient resources to address their backlogs of unprocessed and un- or under-described archival collections. Related to issues of staffing, many organizations rely on contingent labor, student workers, volunteers, or other temporary or non-specialist staff who do not work on a full-time basis. Participants described this practice as detrimental to progress because the new staff needs to be repeatedly trained, volunteers and short-term staff only can be assigned limited types of work, or work needs to be shifted back to full-time staff when temporary positions end.

Archivists are challenged to find both staffing and budgetary resources to move descriptive work forward, especially when balanced against other competing priorities at their institution.

Information technology (IT) resources, or lack thereof, were another key challenge identified in the focus group interviews. Archivists discussed a lack of agency, support, or control over IT services, software, and systems. Archives are highly reliant on specialized systems, and often are dependent on others outside of their department or unit to make purchasing or licensing decisions, to implement and maintain software, and to provide electronic storage solutions.

INCENTIVES TO PARTICIPATING IN AGGREGATION

The top factors cited by the focus group interview participants as a motivator or perceived benefit of participating in an archival aggregation is the potential for increased visibility or awareness of collections in their care. They expressed that a national aggregation would make collections discoverable in a way that would not be possible for an institution to do on its own. This visibility was discussed as beneficial for increasing knowledge of and access to their own collections, especially for researchers who may not know about the repository.

Key to increasing visibility and discovery of archival materials was the ability of a national aggregation to uncover connections among collections held at different institutions. These connections could be around collection creators, other people or organizations, topics, and places, and would be useful both to researchers and to archivists doing reference and collection development work. Related to this increased visibility, some discussed the value of the aggregator to quantify traffic and engagement through analytics or reporting.

Another important aspect of the discovery system cited by participants is support for digital objects to be integrated into the aggregation discovery interface. Other features mentioned by participants included support for advanced search or faceting of results, including the ability to filter or narrow a search to a specific repository or geographic location, and support for a simple and direct way to contact an archive once an end user has located a collection of interest. Some participants said that they expected the system to support discovery by being easy for users to understand, using terms such as “simple” and “intuitive,” as well as expressing that the system should be designed with end

users in mind. Participants tied ease-of-use expectations to their ability to serve their mission and to advocate for participation in an aggregation system. There was an expectation that the aggregation would put some effort into marketing itself and work with search engines to surface collection descriptions represented in the aggregation.

While focus group interview participants saw a range of exciting possibilities in participating in an aggregation, they also made it clear that contributing their descriptive records needed to be easy and must not require redundant work. Participants expressed a desire for the aggregation to harness the other systems being used by archivists, including automatic uploads, crawling, or harvesting existing records, or working with the CMS directly. Several participants discussed the need for participating in an archival aggregation to present a “low barrier to entry.” This phrase was used to articulate concerns both around ease of use and cost. Participants described a need for the systems to be easy for archivists to use, either because of limited time due to many competing priorities, or due to a lack of technical skills. Some speakers linked ease to automated harvesting while others imagined a simple way to input data via a record creation tool provided by the aggregator.

Another aspect that might motivate archivists to contribute to an aggregation is if it supports participation from smaller or less well-funded institutions, such as community archives and historical societies. Nineteen participants discussed the need for a more inclusive aggregation, one that was not built to privilege and benefit only large, well-funded, and prestigious institutions. In addition to being a destination for users, focus group interview participants saw a national aggregation program as being able to foster a sense of community and connection and to serve as a locus of peer learning for archival practitioners.

DISINCENTIVES TO PARTICIPATING IN AGGREGATION

The primary barriers to contribution to archival aggregations are a range of difficulties in contributing and maintaining records with the aggregator. The top barrier to contribution is having to comply with the aggregator’s data requirements. These might be related to the format or structure of the archival description data files, and often required participants to recreate or revise records for the aggregator system that they already had created in their local system.

Participants also discussed how difficult it can be to add, edit, or delete records that are represented in an aggregation. Some participants had difficulty confirming that their records had been updated or deleted and some had a more basic problem of difficulty adding collection description into aggregation systems. Participants were frustrated by both their lack of control over their own data in these situations and the repetitive work and wasted time caused by cumbersome processes.

Another potential disincentive described by those interviewed was a concern for both the investment of time and resources required to participate in aggregation and a need to show the impact of that investment. Eight focus group interview participants discussed cost as a potential barrier to participation, either for their own institution or as a perceived barrier for other institutions. Four participants talked about the need to get buy-in from decision makers. In many institutions, the archivist is not a high-level resource allocator and might have to convince someone else that it is worthwhile to participate in aggregation.

Aggregation contributor research: descriptive data analysis

The research questions guiding the project's metadata analysis focused on determining the structure and extent of consistency across the body of EAD encoded records found in current aggregations of archival description. It also analyzed how that data might support some of the user needs identified in the pop-up survey and semi-structured individual interviews with users of archival aggregation.

DATA STRUCTURE, CONSISTENCY, AND ABILITY TO SUPPORT DISCOVERY

Examining the data quality to understand the structure and consistency across EAD finding aid data is of primary importance for the NAFAN project; this work helps to establish a baseline for what EAD flowing into NAFAN would look like. In undertaking this work, OCLC approached EAD tag and attribute usage from a discovery perspective, using a framework that identified five high-level features that often are present in archival discovery systems: search, browse, results display, sort, and facet.¹⁹ The desire for this functionality was echoed in the semi-structured interviews with users. OCLC's analysis examined the EAD elements and attributes that, if present, could be accessed, indexed, and displayed to facilitate these high-level discovery features.

OCLC assigned threshold levels to better characterize the usage findings:

- Complete: 96–100% usage
- High: 81–95% usage
- Medium: 51–80% usage
- Low: 0–50% usage

The results of the usage analysis are decidedly mixed. Some important elements are used at a high or complete threshold, but many elements needed for discovery interfaces are at medium or low use. Only two elements attained complete usage, collection titles at 99.98% and repository information at 99.68%. Dates, which could potentially be utilized as search terms, or leveraged for browsing, sorting, or faceting, were used at 81.89%. Scope and content notes, abstract notes, and controlled vocabulary elements as a category were the only other elements that achieved the high use threshold. Individually, controlled access elements largely fell in the low-use category, with only subjects (75.1%) and personal names (54.34%) falling in the medium-use category. See table 1 for further detail. It is notable that geographic names only achieved low usage at 34.78%, as being able to search or facet by geography was a functionality specified by multiple individual interview participants.

In addition to the general examination of data structure and consistency, OCLC conducted a focused examination to inform the NAFAN project goal to define a subset of metadata fields that will be required of all records to be added to the system or a NAFAN minimum viable descriptive record. *Describing Archives: A Content Standard* (DACS) is the widely adopted, Society of American Archivists-endorsed content standard for archival description in the United States.²⁰ DACS specifies required and suggested content for descriptive records to be compliant with the standard, and many focus group participants described creating DACS-compliant description. The DACS single-level optimum record is a commonly used level of description across many archives and is therefore a good proxy for investigating what might be required by NAFAN for its minimal viable descriptive record.

Our analysis indicates that extant archival description is in good alignment with the element requirements of the DACS standard, with all but one of the elements included in a DACS single-level optimum record having complete or high use in the data set. The element that only attained a medium usage threshold was bioghist, used for the Biographic or Historical Note for a collection, which was used 72% of the time. In the discovery environment, data in this field are of the most use for keyword searching. As the bioghist element typically contains an unstructured note, it cannot be used to build browse, facet, or sort features and so its lower usage would have limited impact on system functionality.

SURFACING CONNECTIONS BETWEEN COLLECTIONS

In interviews with both archivists and end users, it became clear that a key value that NAFAN can bring is surfacing relationships between archival collections held at different institutions. These relationships could be identifying collections on the same subject, collocating the writings of a single person, or showing relationships across collections by identifying correspondents. Leveraging controlled vocabularies within the data is an obvious way to build out such functionality.

The elements used in EAD to control for names of people, families, organizations, subjects, and places are present in 85% of the data set at both the origination and archdesc levels. Individual elements are used with less frequency. For example, geographic names are used in only 34.78% of the data, and subjects in 75.41%. See table 1 for further detail.

TABLE 1. Controlled vocabulary usage in EAD corpus

| Discovery category | Ead element | Percentage of use |
|---|---------------|-------------------|
| Dates | unitdate | 81.89 |
| Origination and content tags | origination | 85.30 |
| | corpname | 25.25 |
| | famname | 01.62 |
| | name | 00.25 |
| | persname | 54.34 |
| Archdesc/controlaccess and content tags | controlaccess | 85.05 |
| | corpname | 41.84 |
| | famname | 06.09 |
| | function | 00.66 |
| | geogname | 34.78 |
| | name | 00.04 |
| | occupation | 06.97 |
| | persname | 43.98 |
| subject | 75.41 | |

An examination of personal name elements within the NAFAN EAD corpus gives a sense of the challenges of leveraging this controlled vocabulary data within an aggregation system. Personal name elements in the aggregation are a sprawling body of data and will require substantial resource commitments to establish their identity. Only 45% of data in personal name elements occurred five or more times across the corpus of 145,000 finding aids; the remaining 55% represent a long tail of infrequently occurring personal names, some of which include too little data to support effective reconciliation.

Establishing the source of a controlled heading can improve the efficiency and accuracy of reconciling headings with a controlled vocabulary. Inclusion of authority file numbers is infrequent, but identification of controlled vocabulary sources is more common. In the extraction of 1,092,209 personal name elements, 94,365 (8.6%) included an authority file number attribute value while 595,048 (54.5%) included a source attribute value. OCLC's analysis found that clustering personal name elements could provide a path to enriching finding aids with controlled vocabulary identifiers and amplify the effect of some finding aids that used the source and authority file number attributes. A combination of automated and manual reconciliation of personal names to a controlled vocabulary could further enrich the clustered elements.

ACCESS TO DIGITAL CONTENT

Access to digital content was a key desire expressed by users in the pop-up survey and individual interviews, as well as by archivists in the focus group interviews. A possible avenue for presenting digital content in a finding aid aggregation would be to provide a filter or search option that would limit results to those that include online content, or perhaps signal in a result set which items have online content associated with them. Some EAD attributes can be used to associate an element with an external resource as a link, and this can serve as an indicator that there is associated online material.

Our analysis of the external links present in the NAFAN data corpus found that the overall average number of links to external content per finding aid is 4.13. The number of links that are referenced by more than one finding aid for each NAFAN participant varies considerably, with some having many and others having relatively few. The quality of these links was quite high; for nine of the 11 aggregators with links to digital content, 85%–95% of links were viable. The media type for external resources was also analyzed. The research team found that 76% of links are to HTML resources, which in many cases may be used to provide a framework for presenting content which may be an image, streaming media, or an embedded PDF document. An in-depth review of the HTML resources was out of scope of this investigation and would be required to ascertain a fuller picture of linked media resources.

CONNECTING AGGREGATION USERS TO HOLDING ARCHIVES

The discovery system that the NAFAN project envisions will need to connect users with archive staff who can assist with on-site access and other research needs by supplying reliable contact information for institutions with holdings represented in the aggregation.

OCLC examined how widely the address element is used in the NAFAN EAD corpus and found that EAD address values could be a starting point for gathering data for a repository registry, as at least one address can be found for 93% of the 741 repositories represented in the data set. Some remediation of this data would also be required. For example, OCLC's analysis found multiple instances of more than one unique address for the same repository. Address element values include valuable information but important contact details such as email address and website URL are not comprehensive across the data set. Text mining of address values in OpenRefine indicates that the

EAD XML also can supply phone or fax numbers (in 96% of addresses), email addresses (66%), and website URLs (56%). Only two repositories used an entity reference to store address information instead of including it directly in the EAD file.

Interviews with archival users indicate that they value being able to easily find information about how to access archival collections. They want information on terms of use to understand availability and plan research trips, and terms of reuse for understanding if they can obtain a reproduction for their use in a publication or project.

Our analysis of the elements that typically communicate this information in the EAD found that this data may be challenging to present to users in consistent ways. For example, varying encoding practices may present challenges for machine interactions with the element values. Even when the `accessrestrict` field was used to indicate that there are no restrictions on access, the EAD data presented 40 distinct string patterns for categorizing those statements. The brief unstructured text found in these elements did not lend itself to out-of-the-box analysis using Natural Language Processing and Named Entity Recognition software. Information alerting users that collections are stored off-site and require an advance request for access was found variously in four different elements (`accessrestrict`, `userrestrict`, `physloc`, and `phystech`). Such issues may challenge NAFAN to offer a consistent user experience and potential functionality such as limiting search results by access or use restrictions.

Recommendations: Applying Research Insights to the NAFAN Project

Findings across these research activities highlight a variety of concerns and potential directions for a future National Finding Aid Network. The following sections consider these findings, along with larger NAFAN project goals and principles,²¹ to suggest functional and programmatic considerations to be taken into account in future phases of work.

Value proposition for a national archival aggregation

The report from the 2018–2019 NAFAN planning grant states that “aggregators promote broader visibility of their contributing institutions’ finding aids, primarily by facilitating search engine exposure,” and that “aggregators strongly perceive a continued value in aggregation” but do not have formal evidence to support this assertion.²² A core principle guiding work on this phase of work on NAFAN is that the project team must gather evidence to better understand and articulate the value proposition that a national aggregation can provide.

One way to think about value proposition is to identify the problems a system is trying to solve, and then articulate how the system will address them. OCLC’s research identifies challenges faced by both researchers and archivists that a national aggregation can address, as well as alignment between the needs of these two groups.

ADDRESS DISCOVERY CHALLENGES

The central challenge surfaced from both researchers and archives is discoverability of archival collections on the web. Researchers in individual interviews described being frustrated by having to use siloed discovery systems and being challenged by the significant amount of time and effort they must expend in identifying archival material of relevance to their research. Archivists described being challenged to make their collections visible to a broad audience of researchers. Both want an easy-to-use system that provides comprehensive discovery of archival collections across many institutions. Similarly, both groups of stakeholders were concerned not just with discovering individual collections, but also surfacing connections between collections.

Archivists highlighted the importance of these connections given the deep interrelatedness of distinct archival collections held across many institutions and because the sometimes-fractured nature of archival collecting means that not all archival material by a single creator may end up in the same archive. Users desire systems that provide connections across collections, institutions, and topics. A national aggregation can provide important value by making archival collections more visible and discoverable on the web and making connections between collections much more easily visible to users.

ADDRESS ACCESS CHALLENGES

While aggregators are traditionally considered discovery systems, research findings indicate that NAFAN can create value by supporting access as well as discovery. End users described a desire to be able to access archival materials once they had discovered them and noted the ways an aggregation system can support or frustrate that process. The NAFAN platform can provide value by understanding how the system can support access needs, planning for archival research trips, and contacting individual archives. Users also noted barriers to doing in-person archival research, especially when it requires traveling to an archive. A national aggregation can provide value by helping researchers discover collections held geographically close to them that they are more likely to be able to easily access. Both pop-up survey data and interviews with users tell us that online access to digital collections is highly desired by researchers. NAFAN can provide important value to researchers by considering how digital resources will be incorporated into aggregation features and interfaces to allow researchers to access digitized collections more easily.

While aggregators are traditionally considered discovery systems, research findings indicate that NAFAN can create value by supporting access as well as discovery.

ADDRESS ARCHIVISTS' CHALLENGES

Archivists face significant barriers and challenges in creating and publishing archival description. Inefficiencies in participating in current archival aggregation systems can exacerbate these issues, creating additional or redundant work for archivists and preventing participation in aggregation. Archivists in the focus group interviews voiced frustrations with inefficient current systems for uploading, editing, and deleting records in aggregations. They also described struggles to create

the EAD encoded finding aids required by most aggregators, or cumbersome processes to alter existing EAD to meet aggregator data requirements, and a lack of resources to get their description in non-EAD formats into aggregators. A national archival aggregation could make important interventions to address these challenges and make creating and sharing archival description easier, make archival collections more visible, and bring benefit to individual institutions and the archival community.

Given the limited resources available to devote to creating descriptive data in archives—and the NAFAN EAD data analysis findings—it is clear there are numerous opportunities to enhance and enrich data in service of supporting a robust and rich discovery experience, as well as considerations for opportunities to return enriched data to contributors. There are many activities that would add value in a discovery environment, such as linking to vocabularies, disambiguating names, and cleaning up dates. The NAFAN project may be able to provide value through data remediation or enrichment efforts that many archives could not support on their own. The next phases of work on NAFAN should consider which efforts would offer most value to the aggregation users and participants and where in the data creation chain enrichment might occur.

ADDRESS INEQUITY IN THE ARCHIVAL LANDSCAPE

Our research identifies that there is significant unevenness in the resources available to both archives and archival researchers, with real impacts on the historical record, scholarship, and knowledge access and dissemination.

Gaps and lacunae in the archival record have been a concern of archivists and scholars alike for some time now. This concern usually centers on what does not make it into the archive, but it also can extend to the difficulty of finding what does exist in archives, especially archives with limited resources to steward the collections in their care. Both end users and archivists were concerned with the visibility of the collections held by small and underfunded archives. Users fear that they are not finding collections held in these institutions; archivists are concerned with the diversity and inclusiveness of the archival record available to researchers and the ability of colleagues in resource-strapped institutions to create and share archival descriptions online.

These concerns align with the NAFAN project guiding principle that “the network must support meaningful, inclusive, and low-barrier pathways to participation by cultural heritage institutions across the United States.”²³ Focus group interview participants offered solutions to address the needs of smaller institutions and make it easier for them to describe the collections in their care and participate in a national aggregation. In some cases, participants voicing these concerns were affiliated with small institutions themselves, in others, they were not. This may point to a need for further investigation into the needs of small and under-resourced institutions to ensure that the NAFAN project addresses true rather than perceived needs. Supporting the needs of these types of institutions to participate in aggregation can help to make their collections more visible, and in turn, impact the kind of research that can be done and what is passed on as public knowledge.

Our research with end users also identified inequity related to who can access archival collections and how easily they can do so. Resource limitations and caregiving responsibilities impacted users’ ability to do in-person research. Some users were limited to working only with digital collections that they could access online. Users working in academic and cultural heritage settings cited using personal and professional networks to navigate the sometimes-opaque ways that archives operate or to avoid barriers other researchers had to routinely deal with. A national aggregation could make an important intervention in this uneven landscape of access by addressing the information needs of a broad range of users.

Potential challenges to sustainability

A guiding principle for the NAFAN project has been “we must design our work with sustainability in mind from the outset.”²⁴ OCLC’s research points to two major areas of mismatch between resources and expectations that may be potential challenges for sustainability.

First, there is a tension between attitudes around aggregation data requirements, users’ discovery desires, and archivists’ discovery expectations. Focus group interview participants described realities related to their archival description resources and practices that do not align easily with the discovery functionality both users and archivists described as desirable. The top barrier that focus group interview participants identified to contributing to aggregations is having to comply with the aggregator’s data requirements, indicating both the data format and data structure requirements as challenges. At the same time, both archivists and end users expressed desires and expectations related to discovery that require structured data. Discovery benefits were archivists’ most-cited incentives to participate in aggregation. Primary among them was a desire to make it easier to discover connections between collections held at different repositories, based on people, organizations, places, and topical subjects. While a basic attempt at this might be accomplished with keyword searching, structured data in the form of authorities or entities would support more sophisticated and reliable functionality to illuminate connections between collections. Participants expressed a strong desire for digital objects to be seamlessly supported within the aggregation, and the utility of advanced search, filtering and faceting, and narrowing searches to specific geographic areas, all of which would rely on structured data.

An aggregator discovery platform that features a simple, intuitive, and easy-to-use interface was another key expectation of both researchers and archivists, a desire archivists tied both to serving users and advocating to administrators for participating in aggregation. This functionality will be difficult to achieve with variable data structure and formats. Indexing a combination of structured and unstructured data may create challenges for search weighting and relevancy ranking. It also will create challenges for how to display search results and for creating understandable paths for users navigating from search results to finding aids and other content pages.

Second, but no less important, is the tension between archivists’ expectations for a low- or no-cost service and the level of desired contributor support and aggregation features. Focus group interview participants expressed a desire that, on the one hand, an archival aggregation should support low barrier participation to a range of institutions and robust discovery features while at the same time being low- or no-cost. The desired features and community support identified by participants would require a significant level of monetary support, and it is likely that contributors would need to bear a major portion of ongoing costs. Discussion of costs in the focus group interviews was abstract and occurred organically during the interviews; the interview protocol did not include any questions directly about cost or requests to react to examples of specific fees. Further investigation is required to assess what costs the archival community are willing to bear to participate and how the market might be segmented to shape a fee structure that is equitable across many types and sizes of institutions.

CONCLUSION

Our findings indicate that there is significant value to be drawn from a national aggregation of archival description, and there are significant challenges to overcome to build the community of participation a national finding aid aggregation will require to be sustainable. The first era of work on finding aid aggregation focused on technical standards implementation; this next era must think more expansively. OCLC's findings point to needs for design, research and testing, and community and user support that will require a holistic conception of the NAFAN system and its resourcing. Concrete guidance can be drawn from this research for prioritization of the next phases of the NAFAN project, not just for technical requirements but also for design and usability, governance, marketing, community building, and further research.

ACKNOWLEDGMENTS

This project would not have been possible without funding from the Institute for Library and Museum Services. The Building a National Finding Aid Network (NAFAN), was a two-year, IMLS-funded (LG-246349-OLS-20) research and demonstration project led by the California Digital Library. OCLC provided significant co-investment for the qualitative and quantitative research that contributed to the success of this effort.

Significant thanks are due to the California Digital Library for their overall coordination of the NAFAN project, and to project partners and collaborators at Shift Collective, University of Virginia Libraries, and Chain Bridge Group. Our gratitude also to members of the project's Aggregator Partners Working Group for their engagement throughout the project, and for providing EAD files for our analysis and hosting the pop-up survey on their websites.

The NAFAN Research Advisory Working Group gave invaluable guidance and perspectives, asking smart questions, helping to review and shape draft project findings and reports, and posing additional areas for inquiry. Our work benefited from the blend of rich professional backgrounds and institutional types represented in this group.

- Hillel Arnold, Assistant Director for Digital Programs, Rockefeller Archive Center
- Rachael Hu, User Experience Manager, California Digital Library
- Bergis Jules, Shift Collective
- Holly Mengel, Archivist, University of Pennsylvania, Kislak Center for Special Collections, Rare Books and Manuscripts
- Derek Mosley, Archivist/Division Manager, Auburn Avenue Research Library on African American Culture and History, Fulton County Library System
- Molly Bruce Patterson, formerly Digital Archivist & Special Collections Librarian, Rhode Island College
- Ricky Punzalan, Associate Professor, University of Michigan
- William Stingone, Rabinowitz Director for Preservation and Collections Processing, New York Public Library
- Lydia Tang, Outreach and Engagement Coordinator, Lyrasis
- Adrian Turner, Senior Product Manager, Archives, California Digital Library
- Rachel Walton, Digital Archivist, Rollins College

We are grateful to all of the people who participated in the pop-up survey, focus group interviews, and individual interviews for sharing their valuable experience and insight with us.

OCLC colleagues contributed to this work. Our gratitude to Janet Mason (Executive Assistant to the Vice President) for her adept management of all the communication and scheduling that made the focus group and individual interviews possible. Thanks also to Brian Lavoie (Senior Research Scientist), Ixchel Faniel (Senior Research Scientist), and Kathryn Stine (formerly Senior Program & Engagement Officer) for pitching in to help us fill last minute gaps during the individual interview process. Thanks to Devon Smith (Technical Manager) and Chris Cyr (formerly Associated Research

Scientist, OCLC) for their contributions to our EAD and other data analysis. We also recognize Kendra Morgan (Senior Program Manager) for her superior knowledge of grant administration and support on this project.

No OCLC Research output would be possible without the efforts of the OCLC Research Communications Team who bring patience, persistence, and creativity to their work. Erica Melko (Senior Communications Coordinator/Editor) makes everything we write better by smoothing and adjusting text and ensuring that our words connect with the intended audience. Jeanette McNicol brings her expertise in layout, displaying gusto and courage in the face of tables. JD Shipengrover (Information Architect) uses her skills as a graphic designer to create meaningful and effective graphics that help readers connect with data.

Finally, this work was made possible by the senior leadership of OCLC. The authors would like to thank Rachel Frick, Executive Director, Research Partnerships and Engagement and Chip Nilges, Vice President, Business Development, Membership and Research, for their ongoing support, and Lorcan Dempsey, formerly Vice President, Membership and Research, for his support for OCLC taking on this project.

NOTES

1. Allison-Bunnell, Jodi, and Adrian Turner (ed.) 2019. *Finding Aid Aggregation at a Crossroads*. UC Office of the President: California Digital Library.
<https://escholarship.org/uc/item/5sp13112>;

At the time of this report, 16 regional aggregators existed in the US. Since then, two aggregators have gone defunct, bringing the current number to 14.

2. Langa, Lesley A., Chela Scott Weber, and Lynn Silipigni Connaway. 2023. *Pop-Up Survey: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/qfjb-h531>;

Washburn, Bruce, Merrilee Proffitt, and Chela Scott Weber. 2023. *EAD Analysis: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research.
<https://doi.org/10.25333/atn7-qq32>;

Weber, Chela Scott, Itza A. Carbajal, Lesley A. Langa, Lynn Silipigni Connaway, Brooke Doyle, Brittany Brannon, and Merrilee Proffitt. 2023. *User Interviews: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research.
<https://doi.org/10.25333/xq53-yv76>;

Weber, Chela Scott, Merrilee Proffitt, Lesley A. Langa, Lynn Silipigni Connaway, Brittany Brannon, and Brooke Doyle. 2023. *Focus Group Interviews: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research.
<https://doi.org/10.25333/szsb-9755>.

3. Turner, Adrian, Jodi Allison-Bunnell, Bradley Daigle, Elizabeth Dunham, Karen Eberhart, and Brian Tingle. 2019. *Toward a National Archival Finding Aid Network—From Planning Initiative to Project and Program: An Action Plan*. UC Office of the President: California Digital Library.
<https://escholarship.org/uc/item/3g64n275>.

4. Allison-Bunnell, Jodi, and Adrian Turner (ed.) 2019. *Finding Aid Aggregation at a Crossroads*. UC Office of the President: California Digital Library.
<https://escholarship.org/uc/item/5sp13112>.

Turner, et al. *National Archival Finding Aid Network*, (see n. 3).

5. Rhee, Hea Lim. 2015. "Reflections on Archival User Studies." *Reference & User Services Quarterly* 54, no. 4 (summer): 29–42. <https://doi.org/10.5860/rusq.54n4.29>.

6. Ibid.

7. See, for example:

ArchivesSpace. 2016. "User Personas." Created by Mark Custer. Updated 23 July 2016. <https://archivesspace.atlassian.net/wiki/spaces/ADC/pages/66355248/User+Personas>;

Stanford University Libraries, Georgia Tech, and University of Michigan. 2017. "ArcLight Personas." Stanford Digital Repository. <https://purl.stanford.edu/hk349dn1751>;

Rockefeller Archive Center. "Project Electron Personas; 2017-19." Personas. <https://github.com/RockefellerArchiveCenter/personas/tree/base/project-electron-personas>.
8. Duff, Wendy M. 2005. "Working as Independently as Possible: Historians and Genealogists Meet the Archival Finding Aid," in *The Power and the Passion of Archives: A Festschrift in Honour of Kent Haworth*, edited by Reuben Ware, Marion Beyea, and Cheryl Avery. Ottawa: Association of Canadian Archivists.
9. Allison-Bunnell, Jodi, and Adrian Turner (ed.) 2019. *Finding Aid Aggregation at a Crossroads*. UC Office of the President: California Digital Library. <https://escholarship.org/uc/item/5sp13112>.
10. See, for example:

Tatem, Jill. 1998. "EAD: Obstacles to Implementation, Opportunities for Understanding." *Archival Issues* 23 (2): 155–70. <http://digital.library.wisc.edu/1793/45878>;

Yaco, Sonia. 2008. "It's Complicated: Barriers to EAD Implementation." *The American Archivist* 71 (2): 456–75. <https://doi.org/10.17723/aarc.71.2.678t26623402p552>;

Roth, James. 2001. "Serving Up EAD: An Exploratory Study on the Deployment and Utilization of Encoded Archival Description Finding Aids." *The American Archivist* 64 (2): 214–37. <https://doi.org/10.17723/aarc.64.2.e687471v304k0u66>;

Combs, Michele, M.A. Matienzo, Merrilee Proffitt, and Lisa Spiro. 2010. *Over, Under, Around, and Through: Getting Around Barriers to EAD Implementation*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/vsd3-wm91>.
11. Wisser, Katherine, and Jackie Dean. 2013. "EAD Tag Usage: Community Analysis of the Use of Encoded Archival Description Elements." *The American Archivist* 76 (2): 542–66. <https://doi.org/10.17723/aarc.76.2.x4h78gx76780q072>;

Bron, Marc, Merrilee Proffitt, and Bruce Washburn. 2013. "Thresholds for Discovery: EAD Tag Analysis in ArchiveGrid, and Implications for Discovery Systems." *The Code4Lib Journal*, no. 22 (October). <http://journal.code4lib.org/articles/8956>.
12. Connaway, Lynn Silipigni, and Marie L. Radford. 2021. *Research Methods in Library and Information Science*, 59. Seventh edition. Library and Information Science Text Series. Santa Barbara, California: Libraries Unlimited.
13. Ibid., 343-368.

14. Langa, Lesley A., Chela Scott Weber, and Lynn Silipigni Connaway. 2023. *Pop-Up Survey: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/qfjb-h531>;
- Washburn, Bruce, Merrilee Proffitt, and Chela Scott Weber. 2023. *EAD Analysis: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/atn7-qq32>;
- Weber, Chela Scott, Itza A. Carbajal, Lesley A. Langa, Lynn Silipigni Connaway, Brooke Doyle, Brittany Brannon, and Merrilee Proffitt. 2023. *User Interviews: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/xq53-yv76>;
- Weber, Chela Scott, Merrilee Proffitt, Lesley A. Langa, Lynn Silipigni Connaway, Brittany Brannon, and Brooke Doyle. 2023. *Focus Group Interviews: Findings from the Building a National Finding Aid Network Project*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/szsb-9755>.
15. For more information, see:
- ArchivesSpace. 2016. "User Personas." Created by Mark Custer. Updated 23 July 2016. <https://archivesspace.atlassian.net/wiki/spaces/ADC/pages/66355248/User+Personas>;
- Stanford University Libraries, Georgia Tech, and University of Michigan. 2017. "ArcLight Personas." Stanford Digital Repository. <https://purl.stanford.edu/hk349dn1751>;
- Rockefeller Archive Center. "Project Electron Personas; 2017-19." Personas. <https://github.com/RockefellerArchiveCenter/personas/tree/base/project-electron-personas>.
16. For example, the ArcLight Personas represent this as a "Community Researcher" and the Project Electron personas represent this as a "Personal Interest Researcher." See:
- Stanford University Libraries, Georgia Tech, and University of Michigan. 2017. "ArcLight Personas." Stanford Digital Repository. <https://purl.stanford.edu/hk349dn1751>;
- Rockefeller Archive Center. "Project Electron Personas; 2017-19." Personas. <https://github.com/RockefellerArchiveCenter/personas/tree/base/project-electron-personas>.
17. For more information, see:
- ArchivesSpace. 2016. "User Personas." Created by Mark Custer. Updated 23 July 2016. <https://archivesspace.atlassian.net/wiki/spaces/ADC/pages/66355248/User+Personas>;
- Stanford University Libraries, Georgia Tech, and University of Michigan. 2017. "ArcLight Personas." Stanford Digital Repository. <https://purl.stanford.edu/hk349dn1751>;
- Rockefeller Archive Center. "Project Electron Personas; 2017-19." Personas. <https://github.com/RockefellerArchiveCenter/personas/tree/base/project-electron-personas>.
18. Legacy description is often characterized as outdated archival description that may be incomplete or inaccurate, may not have been reviewed or updated in many years, may not be compliant with current descriptive standards, and/or may be in paper-based or older electronic formats that cannot be easily used by modern systems.

19. This framework was originally developed in earlier work coming out of OCLC Research. See:
Bron, Marc, Merrilee Proffitt, and Bruce Washburn. 2013. "Thresholds for Discovery: EAD Tag Analysis in ArchiveGrid, and Implications for Discovery Systems." *The Code4Lib Journal*, no. 22 (October). <http://journal.code4lib.org/articles/8956>.
20. Society of American Archivists' Technical Subcommittee on Describing Archives: A Content Standard (TS-DACS). 2022. "Describing Archives: A Content Standard—Version 2022." <https://saa-ts-dacs.github.io/>.
21. Allison-Bunnell, Jodi, and Adrian Turner (ed.) 2019. *Finding Aid Aggregation at a Crossroads*. UC Office of the President: California Digital Library. <https://escholarship.org/uc/item/5sp13112>;

Turner, Adrian. 2021. "Our Goals, Values, and Principles." *Building a National Finding Aid Network*. Confluence, University of California Office of the President. Created by dbartlin 10 October 2018. Last modified 12 March 2021. <https://confluence.ucop.edu/display/NAFAN/Our+Goals%2C+Values%2C+and+Principles>.
22. Allison-Bunnell, Jodi, and Adrian Turner (ed.) 2019. *Finding Aid Aggregation at a Crossroads*. UC Office of the President: California Digital Library. <https://escholarship.org/uc/item/5sp13112>;
23. Turner, Adrian. 2021. "Our Goals, Values, and Principles." *Building a National Finding Aid Network*. Confluence, University of California Office of the President. Created by dbartlin 10 October 2018. Last modified 12 March 2021. <https://confluence.ucop.edu/display/NAFAN/Our+Goals%2C+Values%2C+and+Principles>.
24. Ibid.