

Statement of National Need

Lack of discoverability, representation and resources

Archival records are primarily discoverable via finding aids, which provide descriptions and summaries of the contents of individual archival collections.¹ These critical scholarly resources enable students, teachers, genealogists, historians, and a broad array of researchers to discover, locate, identify, and understand archival resources, especially those in print format. Unfortunately, finding aids themselves are often difficult to find; many institutions do not publish finding aids online or, if they do, display these materials on their own platforms with varying success in making them visible to a broad community or to search engines. As a result, researchers face enormous barriers in locating relevant materials across the vast, distributed, and unevenly supported field of cultural heritage institutions, rendering much of that stewarded content essentially invisible.

Finding aid aggregators, that collect and host finding aids on behalf of multiple institutions, have historically made strong inroads towards addressing these discoverability challenges.² Along with providing persistent hosting of finding aids, the aggregators provide additional core services (e.g. free tools, training, and support) to help institutions create and share finding aids in standard formats. Many aggregators serve as the de facto or primary publication venue for those institutions that lack the capacity to maintain their own publication system. In all cases, aggregators provide additional visibility for the finding aids they collect.³

Over the past two decades, aggregators have formed at the state and regional level, with significant early investments by institutions, states, and funders, and now number sixteen across the country.⁴ Despite this substantial base, however, researchers still lack broad and persistent access to archival materials via finding aids: not all US regions have a finding aid aggregator, and the aggregators that do exist maintain siloed platforms that display finding aids from institutions only within a delimited service area. This geographically restrictive demarcation of holdings information interferes with a researcher's ability to comprehensively explore archival resources related by more meaningful dimensions, such as provenance or subject matter.

Compounding this problem, thousands of cultural heritage institutions still are not represented within extant aggregations. Many organizations have finding aids and collection descriptions in formats unsupported by most aggregators, such as spreadsheets, MS Word, HTML, and PDF. As a result, countless small, lesser-resourced community-based archives are invisible -- and the voices documented in their collections are poorly represented in the historical record.

Complementary platforms such as SNAC and ArchiveGrid⁵ also play a role in the national landscape, leveraging and indexing finding aid and other data gathered from aggregators and individual institutions -- and serving as an additional access point for researchers to identify relevant archival holdings. However, SNAC and ArchiveGrid themselves are dependent on the persistent access to finding aids made available by aggregators (and individual institutions). Apart from the significant challenge of providing broader access to distributed archival materials, many aggregators struggle even to find sufficient resources to maintain the status quo, let alone invest in improvements; with aging systems and negligible budget lines, they are fundamentally at risk, which is in turn a major risk for the archival community and researchers around the world. The dissolution of aggregators would spell the loss of many online finding aids (where the aggregator has served as the institution's primary publication platform) and could cost the community both resources and users. Any

¹ Society of American Archivists (SAA) glossary: <https://www2.archivists.org/glossary/terms/f/finding-aid>.

² For a detailed and current profile of finding aid aggregators, see Allison-Bunnell, Jodi (May 2019), "Finding Aid Aggregation at a Crossroads," <https://escholarship.org/uc/item/5sp13112>.

³ See **Appendix E** for testimonials and letters of support from a small but representative group of OAC contributors.

⁴ Archival Resources in Wisconsin; Archives West; Arizona Archives Online (AAO); Black Metropolis Research Consortium (BMRC); Chicago Collections Consortium (CCC); Connecticut's Archives Online (CAO); Empire Archival Discovery Cooperative (EmpireADC); History of Medicine Finding Aids Consortium (HOMFC); OhioLINK Finding Aid Repository; Online Archive of California (OAC); Philadelphia Area Archival Research Portal (PAARP); Rhode Island Archival and Manuscript Collections Online (RIAMCO); Rocky Mountain Online Archive (RMOA); Texas Archival Resources Online (TARO); Virginia Heritage; and University of Nebraska Consortium of Libraries (in development).

⁵ Social Networks and Archival Context Cooperative (SNAC): <https://snaccooperative.org>. ArchiveGrid: <https://www.oclc.org/research/themes/research-collections/archivegrid.html>.

attempts at institution-based solutions would, by definition, result in greater duplication of effort and an even more scattered landscape of finding aids for researchers.

A timely opportunity and a path forward

These challenges, though significant, provide a timely opportunity to rethink aggregation at scale. Our shared imperative is to provide researchers of all types with richer, more comprehensive access to collections held by a diverse array of US cultural heritage institutions and documenting a far broader and representative set of communities, voices, and topics. And we need to do this in a more sustainable way than we have managed in the past, by abandoning siloed, duplicative efforts and collaboratively leveraging a large-scale, national effort.

In 2018-2019, the California Digital Library (CDL) organized an LSTA-funded planning initiative, "Toward a National Archival Finding Aid Network" (NAFAN), convening a symposium of US finding aid aggregators and expert advisors with deep knowledge of organizational development, community engagement, and sustainability⁶ to explore this new aggregation model. Preparatory research and outcomes from the symposium clearly confirm that current aggregations are insufficient in their coverage of archival materials and at-risk in their current siloed contexts.⁷

Symposium participants called for a substantially more inclusive framework that meets the needs of a more diverse set of end users, enables participation by a wider 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. They conceived of this as a national finding aid network that is community-driven, -sustained, and -governed, and formulated an action plan outlining a concrete path toward this set of goals.⁸ Participants identified, as core to this network, a suite of shared infrastructure and services supporting:

- Meaningful, inclusive, and low-barrier pathways to participation by any cultural heritage institution in the US, accepting formats beyond Encoded Archival Description (EAD) and MARC21⁹ and providing tools and support to easily create and publish finding aids.
- An extensive and continually updated registry of institutions, regardless of whether or not they have finding aids, to increase the awareness of organizational profiles and holdings among researchers.¹⁰
- A primary or supplemental publication platform (as needed by the holding institution) with comprehensive aggregation of and persistent access to finding aids, as well as integration with related context and content (e.g., SNAC, DPLA).¹¹

Project category and phase of maturity

Our proposed "Building the Foundation for a National Finding Aid Network" 2020-2022 research and demonstration project aligns with the IMLS project category "National Digital Infrastructures and Initiatives" and the pilot phase of maturity. This nationally-scoped project will identify viable technical solutions with transformative, wide-ranging impact for researchers and contributors from all backgrounds across the country.

Building on existing theory, scholarship, and practice

We are building on prior and current initiatives to promote broader, ethical, and equitable pathways to

⁶ NAFAN project wiki: <http://bit.ly/nafan-project>.

⁷ See Allison-Bunnell, Jodi, (May 2019). See also the notes from the symposium: <https://confluence.ucop.edu/display/NAFAN/Partner+Meetings>.

⁸ Allison-Bunnell, Jodi et al. (October 2019). "Toward a National Archival Finding Aid Network - From Planning Initiative to Project and Program: An Action Plan," <https://bit.ly/action-plan-nafan>.

⁹ See SAA glossary, <https://www2.archivists.org/glossary/terms/e/encoded-archival-description> and <https://www2.archivists.org/glossary/terms/m/marc>.

¹⁰ Such a registry could initially draw on the nearly 5,000 and rapidly growing list of institutions included in SNAC and be integrated with community organizational identifier initiatives, such as the Research Organization Registry (ROR; <https://ror.org/>). It could also leverage the extensive set of repository location information gathered through the RepoData project (<https://github.com/RepoData/RepoData>).

¹¹ Digital Public Library of America (DPLA): <https://dp.la/>.

connect researchers with cultural heritage institutions and the collections they steward. An example is the IMLS-funded 2020 "Lighting the Way: National Forum on Archival Discovery and Delivery" initiative (LTW), the outcomes of which promise to be highly relevant to our project.¹² Our project will substantially update the field's current understanding of researcher needs and their information-seeking behaviors and will also expand knowledge regarding the quality of data needed within finding aids to effectively meet those needs. In a less well explored area, the project will build on the 2018-2019 LSTA-funded NAFAN planning initiative to gain a deeper understanding of factors that hinder or promote participation in aggregations by cultural heritage institutions. This project also addresses challenges faced by aggregators struggling with siloed, unsustainable platforms, a fundamental issue with parallels across the broader scholarly communications landscape.¹³

Finally, our proposed national finding aid network will exist within a complex ecosystem of related platforms such as SNAC, ArchiveGrid, and collection management tools such as ArchivesSpace.¹⁴ No existing system provides the full suite of functionality and breadth of contributor engagement envisioned for this network. SNAC aggregates information about people and organizations that created or are significantly documented in archival collections, but has not addressed the challenge of aggregating finding aids. ArchiveGrid is primarily based on MARC21 data extracted from WorldCat, does not reflect all contributors in the US, and does not provide a persistent finding aid hosting and publication service, making it dependent on the sustainability of existing statewide/regional aggregations and individual cultural heritage institutions' abilities to maintain their finding aids online. ArchivesSpace is a widely-used archival collection management system,¹⁵ but is not designed to serve as a core aggregation, indexing, and publication platform for potentially millions of finding aids. The various players occupying the archival description and access landscape play distinct, intersecting roles, but there has not yet been a concerted effort to rationalize and coordinate activities to develop an overarching collaborative vision. We are therefore intentionally partnering with organizations maintaining key services within the ecosystem--the University of Virginia Library (SNAC), OCLC (ArchiveGrid), and LYRASIS (ArchivesSpace)--to ensure our work is complementary and not competitive.

Project Design

Project goals and assumptions

Our primary goals for this research and demonstration project are to progress towards the goals of equity and sustainability by launching a research agenda to guide the mid- to long-term development of the network, generating community engagement through activities that establish a shared vision for the network, and demonstrating the viability of scaling current siloed aggregation activities to a national level.

Potential risks do exist. A shared solution may not have the flexibility to accommodate all aggregator and individual cultural heritage institution needs, such as distinct branding or unique finding aid formatting preferences. If the sustainability, governance, and service models are not well-tuned to the needs and participatory levels of aggregators and individual institutions, network-level activities will be difficult to maintain. The potential benefits, however, outweigh the risks. By taking action to address the potential dissolution of existing statewide/regional finding aid aggregators over time, we can avoid the possibility of a landscape where individual institutions need to maintain local solutions that serve a smaller subset of researchers, requiring duplication of effort, and representing an incomplete historical record with less-resourced cultural heritage organizations unable to expose their collections.

¹² See <https://library.stanford.edu/projects/lightingtheway>. CDL staff participated in the February 2020 symposium. See also **Appendix D** for a letter of support from the LTW Project Director.

¹³ See Educopia's Mapping Scholarly Communication Infrastructure initiative: <https://educopia.org/mapping-scholarly-communications-infrastructure/>. See also the Invest in Open Infrastructure initiative: <https://investinopen.org>.

¹⁴ ArchivesSpace: <https://archivesspace.org/>.

¹⁵ Spiro, Lisa (2009), "Archival Management Software: A Report for the Council on Library and Information Resources," (Washington, DC: Council on Library and Information Resources), https://www.clir.org/wp-content/uploads/sites/6/spiro_Jan13.pdf.

Project activities and roles

Our project has four core activity areas:

1. **Research:** Initiate research to gain a current understanding of the needs of both researchers and cultural heritage institutions in regard to finding aid aggregation. Evaluate the quality of existing finding aid data at scale in order to scope the network's initial functionality to the specific existing characteristics of the data, and lay the groundwork for iterative data remediation and expanded network features in subsequent phases.
2. **Technical assessment and prototyping:** Conduct technical assessment of existing aggregation systems, develop prototypes to support research activities and demonstrate key network functions, and formulate specific system requirements for a minimum viable instantiation of the network (targeted for a post-2022 implementation phase). Key assessment dimensions will include: open source status and license; adaptability of the code base; size and engagement of the developer community; programming language adoption levels; and compliance with accessibility, standards and universal design principles.
3. **Community engagement:** Cultivate community engagement and broad participation by aggregators, individual cultural heritage institutions, and other stakeholders to support the development and implementation of the network.
4. **Sustainability and governance:** Engage in market research activities to identify viable sustainability and governance models for the network in subsequent phases.

CDL will administer and coordinate the overall project, in collaboration with OCLC and the University of Virginia Library (UVA), and in close partnership with statewide/regional aggregators and LYRISIS. Together, the CDL, OCLC, and UVA bring demonstrated, longstanding experience with multi-organizational approaches to content aggregation at scale, data interchange standards development, and research to advance learning, innovation, and collaboration across the cultural heritage community.¹⁶ CDL will be primarily responsible for overall project administration and coordination, community engagement, sustainability and governance model development, and formulating system requirements for a minimum viable instantiation of the network. CDL will contract with the Chain Bridge Group to conduct sustainability and governance modeling.¹⁷ OCLC will be primarily responsible for conducting and coordinating research activities, and synthesizing and disseminating findings. UVA will be primarily responsible for technical assessment and prototyping activities. See **Appendices A** and **B** for letters of intent and scope of work statements.

CDL, OCLC, and UVA will closely coordinate structured participation of existing aggregators: participating in project consultative structures; supporting research activities; sharing finding aid data for analysis; testing and providing feedback on prototype systems; providing input on system and functional requirements; hosting project meetings; and sharing data to support sustainability and governance modeling. Twelve aggregators are committed to participating in this pilot. See **Appendix C** for letters of support.¹⁸ LYRISIS will provide consultation on future ArchivesSpace integration models.

Project implementation teams and consultative structures

The following project investigatory teams and consultative structures will be used to guide and ensure

¹⁶ The CDL and UVA developed and maintain the OAC and Virginia Heritage, two early statewide finding aid aggregations. CDL and UVA were also involved in the development of SNAC. OCLC maintains ArchiveGrid, lending experience with harvesting approaches to gather finding aid data -- and also brings strong expertise with designing and implementing a range of research activities, with a keen understanding of the archival discovery and delivery domain.

¹⁷ Chain Bridge Group's Managing Partner, Raym Crow, has over 30 years of experience in academic publishing and library information services, specializing in strategic business planning, collective support models for open-access services, and market management. Clients include ORCID, University Press eBook Consortium, University of California Press, and the eLife Initiative.

¹⁸ The University of Florida Libraries, representing the currently-defunct Archives Florida aggregation, is planning to participate. We are in-progress of reaching out to the following five additional aggregators, to secure their participation: Archival Resources in Wisconsin, Black Metropolis Research Consortium (BMRC), OhioLINK Finding Aid Repository, Rocky Mountain Online Archive (RMOA), and the University of Nebraska Consortium of Libraries.

the success of the project, leveraging the use of web conferencing (e.g., Zoom) to host online meetings:¹⁹

- **Project Team:** 7 staff from CDL, OCLC, and UVA. Serves as the primary locus for project activity planning and coordination.
- **Aggregator Partners Working Group:** Approximately 16 senior product managers from regional/statewide aggregator communities, to support and participate in project activities.
- **Technical Working and Advisory Group:** Approximately 10 technical representatives from CDL, UVA, OCLC, LYRASIS, statewide/regional finding aid aggregators, and other potential expert advisors. Provides input and direction on research programming, technical assessment, and development prototyping activities.
- **Research Working and Advisory Group:** Approximately 4 representatives from CDL, UVA, OCLC, and approximately 5 product/service manager representatives from statewide/regional finding aid aggregators and other potential expert advisors. Provides input and support for research activities.
- **Stakeholder Advisory Group:** Approximately 15 senior staff from individual cultural heritage organizations, senior staff from organizations that support related services and programs within the broader scholarly communications ecosystem,²⁰ and researchers. Provides broad community input on project activities, strategies, and direction.

All activities will be conducted in parallel throughout 2020-2021. 2021-2022 will be spent completing a full synthesis of research findings, formalizing system requirements for post-2022 implementation, and establishing initial sustainability and governance models. For additional details on the workplan and personnel/resource allocations, see the **Schedule of Completion** and **Budget Justification**.

Target communities and audiences

Statewide/regional aggregators, cultural heritage institutions, and the broader research community form the project's three core stakeholder sectors and will be engaged through the groups described above, in ways designed specifically to elicit their requirements and input. **Statewide/regional aggregators** will participate through the Aggregator Partners Working Group. Aggregators will also be represented in the task-oriented Technical and/or Research Working and Advisory Groups. Those interested in just providing feedback will do so via the Stakeholder Advisory Group. **Cultural heritage institutions** will serve on the Stakeholder Advisory Group, volunteer as respondents in research efforts, and serve on task-oriented groups of interest. **Researchers** will be included on the Stakeholder Advisory Group and will be key subjects for the research agenda.

The project's group structures will provide explicit paths for participation and continual input, ensuring the project benefits from ongoing dialogue with the full spectrum of stakeholders regarding use cases, priorities, and the degree to which specific project efforts meet them. Project partners will regularly evaluate participation to ensure that: 1) stakeholders across the archival research ecosystem are engaged, from well-established/resourced archives to underserved contributor and end-user communities; 2) feedback from all groups circulate throughout these structures as appropriate; and that 3) documentation regarding work status, end results, next steps, etc. are widely available to all, including the public, through a wiki/project website.

Research questions and theoretical framing

The original NAFAN symposium established a draft research agenda (documented in the action plan²¹) from which we have distilled three areas of investigation, which are foundational for this pilot project:

¹⁹ CDL regularly utilizes Zoom to host and participate in large planning group meetings, such as for DPLA's Network Council, the Library Publishing Coalition, and UC Libraries systemwide initiatives. We will also confer with conveners of large online conferences such as CNI Spring 2020 to ensure use of developing best practices.

²⁰ This includes organizations such as DPLA as well as NARA (which collaborates with UVA to support SNAC's training and educational programs). See **Appendix D** for letters of support.

²¹ See Allison-Bunnell, Jodi et al. (October 2019). "Toward a National Archival Finding Aid Network - From Planning Initiative to Project and Program: An Action Plan," <https://bit.ly/action-plan-nafan>.

- 1) **Researchers:** Who are the current users of finding aid aggregations? Do they align with the persona types and needs identified in recent archival persona work?²² What are the benefits and challenges users face when searching for descriptions of archival materials within finding aid aggregations?

Studies of archival researchers have generally focused on the needs of academics, historians and other humanists.²³ While relatively small, the body of literature investigating a broader cross-section of archival users indicates that different researcher types have significantly different needs.²⁴ To better understand those needs, several institutions have developed personas reflecting a more diverse spectrum of researchers seeking access to collections, ranging from genealogists, local historians, enthusiast researchers, public service librarians, and K-12 educators to journalists, documentary film producers, and authors.²⁵ Based on a synthesis of these personas, our end-user research will include interviews with five types of users: Advanced Academic, Academic, Enthusiast, Professional, and Archivist.²⁶

Additionally, prior investigations have largely focused on the use of and interaction with the collections and interfaces of a single organization's finding aid system.²⁷ Our research will fill a significant gap in the literature: understanding researcher needs as they explore the holdings of multiple institutions within a finding aid aggregation.

- 2) **Cultural heritage institutions:** What are the enabling and constraining factors that influence whether and how institutions describe the archival collections in their care? What are the enabling or constraining factors that influence whether institutions contribute to an aggregation? What value does participation in an aggregation service bring to institutions? Our research will build on information provided by aggregator participants in the NAFAN planning initiative,²⁸ which represents the most current understanding of the interactions between contributors and aggregators. We will extend this work by gathering information from archivists at individual institutions about the factors that hinder or promote the use of finding aid aggregations.
- 3) **Finding aid data quality:** What is the structure and extent of consistency across finding aid data in current aggregations? Can that data support the needs to be identified in the user research phase of the study? If so, how? If not, what are the gaps? Previous research to evaluate the quality of finding aid data at scale focused primarily on EAD tag analysis. We will extend that work by programmatically assessing the presence, quality, and consistency of data structure and content in multiple formats, factoring in collection- and multilevel descriptive elements recommended by US and international archival standards.²⁹ This will lay the groundwork for exploring remediation strategies in future phases. We will gain a current sense of descriptive and encoding practice, which may have shifted significantly

²² Stanford University Libraries, Georgia Tech, and University of Michigan (2017), "ArcLight Personas." Stanford Digital Repository, <https://purl.stanford.edu/hk349dn1751>. See also Rockefeller Archives Center (2019), "Project Electron Personas," <https://github.com/RockefellerArchiveCenter/personas>.

²³ Rhee, H.L. (2015).

²⁴ Duff, Wendy M. (2005), "Working as Independently as Possible: Historians and Genealogists Meet the Archival Finding Aid," *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).

²⁵ Stanford University Libraries, Georgia Tech, and University of Michigan (2017), Rockefeller Archives Center (2019).

²⁶ Stanford University Libraries, Georgia Tech, and University of Michigan. (2017).

²⁷ Rhee, H. L. (2015). "Reflections on Archival User Studies," *Reference & User Services Quarterly*, 54(4), 29–42, <https://doi.org/10.5860/rusq.54n4.29>.

²⁸ Allison-Bunnell, Jodi (May 2019).

²⁹ *General International Standard Archival Description* (ISAD-G) (<https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>) and *Describing Archives: A Content Standard* (<https://saa-ts-dacs.github.io/>). For previous finding aid analyses, see: Bron, M., M. Proffitt, and B. Washburn (2013), "Thresholds for Discovery: EAD Tag Analysis in ArchiveGrid, and Implications for Discovery Systems." *Code4Lib Journal*, no. 22, <http://journal.code4lib.org/articles/8956>. See also Wisser, Katherine, and Jackie Dean (2013), "EAD Tag Usage: Community Analysis of the Use of Encoded Archival Description Elements," *The American Archivist* 76, no. 2: 542–66, <https://doi.org/10.17723/aarc.76.2.x4h78gx76780q072>.

with the broad adoption of tools such as ArchivesSpace, use of the DACS standard, and emphasis on minimal processing.

Research data collection and analysis

1) **Researcher data collection and analysis**

- **Surveys:** A pop-up survey on selected aggregator websites will be used to gather demographic data and information about why and how researchers use finding aid aggregations. Built on SurveyMonkey, it will be deployed until an adequate sample size is reached to support the subsequent interview phase, collecting responses from all willing participants (i.e. using a convenience sampling frame).³⁰ The survey instrument will be pre-tested and subsequently modified as needed prior to implementation (see **Appendix F** for a draft survey instrument).
- **Interviews:** 25 selected survey respondents will be offered incentives to participate in virtual, one-on-one, semi-structured 45-60 minute interviews (a purposive sample constructed of five individuals from each defined user type), to collect more detailed information about why and how individuals use finding aid aggregations. Semi-structured interviews use guided, open-ended, questions asked in the same order during each interview, allowing interviewers to ask probing questions at their discretion, and are a prime methodological strategy used in qualitative research. The interview questions will be pre-tested and revised accordingly (see **Appendix F** for a draft semi-structured interview instrument). The interviews will be digitally recorded and transcribed for coding and analysis.
- **Data analysis:** Survey results will be analyzed using R to generate descriptive statistics and to perform relational tests. Interview responses will be analyzed on the NVivo platform, using a codebook developed from emerging themes. Inter-coder reliability will be calculated to ensure consistency across the coded data set.

2) **Cultural heritage institution data collection and analysis**

- **Surveys:** Participants will be recruited through invitations sent to library, archive, and museum association mailing lists; we will also leverage contacts with regional/statewide aggregators, to identify institutions that are not currently contributing to aggregations. We will use SurveyMonkey to collect responses from willing participants (see **Appendix F** for a draft survey instrument).
- **Focus groups:** Five in-person focus group interviews (approximately 12 participants per group) will be conducted with archivists representing a diverse range of institutions, selected to reflect a variety of sizes, locations, collecting areas, and participation in aggregations. While 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, which is the purpose in this study. Focus group interviews will be digitally recorded and transcribed for analysis.
- **Data analysis:** Responses to the questions will be coded based on emerging themes. Interview responses will be analyzed on the NVivo platform, using a codebook developed from these themes. Inter-coder reliability will be calculated to ensure consistency across the coded data set.

3) **Finding aid data quality collection and analysis**

- **Data gathering and analysis:** A representative sample of finding aid data contributed by aggregators will be programmatically analyzed to: surface the extent of finding aid formats (e.g., EAD, MARC21, HTML, PDF); provide field-level evaluations of structured data and content (when possible, based on the format); note areas of commonality and variances across represented descriptive schemas; indicate the extent of multilingual descriptions; and highlight apparent descriptive trends in community practice. We anticipate using technologies including XPath queries and R to process and evaluate the data. We hypothesize that this analysis will

³⁰ Connaway, Lynn Silipigni, and Marie L. Radford (2017), *Research Methods in Library and Information Science, 6th ed.* (Santa Barbara, CA: Libraries Unlimited).

reveal a heterogeneity of formats; varying quality and extent of data beyond collection-level descriptions; and minimal use of non-English descriptions.

The project will adhere to Institutional Review Board (IRB) requirements for human subjects research. The research team will assist the participating institutions with this process and complete the required training and certificate process. For participating institutions without an IRB process, we will use an external IRB service. Research findings comprising reports, methodologies, and links to datasets (metrics and survey results) will be made openly available via the OCLC Research website.³¹ The findings will be maintained in common data formats. The data collection instruments, codebook for the individual semi-structured interviews, and the focus group interview themes also will be openly available via the OCLC Research website. OCLC Research will coordinate with the CDL to securely store backup copies of all research findings and raw data (consent agreements, interview and focus group transcripts, recordings, etc.) in Merritt, CDL's CoreTrustSeal-certified digital preservation repository.³² These preservation copies will be managed as a "dark archive," only accessible by designated CDL staff solely for disaster-recovery purposes.

Broader dissemination of project outputs

The broader library, archive, and museum community is our primary audience, with a secondary audience found in adjacent communities of practice in the scholarly communication ecosystem (e.g. repository managers, library based publishers, open source technology developers, etc.). We will establish a formal communications/outreach strategy, with supporting collateral, to guide the dissemination of information about the project, research findings, and other deliverables through various channels, including:

- A public-facing project wiki with context and information about the project status, deliverables, and references to all key artifacts.³³
- Open research findings and associated data collection instruments available via the OCLC Research website. We will explore opportunities to publish project information and findings in relevant journals (e.g., SAA's *American Archivist* and *Archival Outlook*, *Code4Lib Journal*) and other appropriate venues (e.g., national and regional library, archive, and museum newsletters).
- Presentations by core Project Team members at four key national conferences (e.g. ALA/RBMS, CNI, SAA). We will coordinate with aggregator partners to disseminate information and findings at regional library, archive, and museum conferences (e.g., within the California context: California Library Association, Society of California Archivists, California Association of Museums).
- Announcements of project milestones, shared through library, archive, and museum mailing lists (e.g., Code4Lib, DLF, EAD, RBMS, SAA Sections), as well as aggregator-specific mailing lists.

As noted under the **National Impact** section of this proposal, the project deliverables (methodologies; data gathering instruments; associated documentation; prototype tools or demonstration systems; and summaries of potential sustainability and governance model options, including descriptions of how those models were arrived at; etc.) will be designed, wherever possible, with an eye toward re-use in other contexts related to developing and sustaining community-based scholarly communications infrastructure.

Success indicators

The Project Team will monitor progress in each activity area as summarized below and elaborated in **Appendix G**. Information will be shared between working and advisory groups to ensure alignment across activity areas, maintain community awareness, and garner input as the need for adjustments arises.

³¹ OCLC Research: <https://www.oclc.org/research/home.html>.

³² Merritt: <https://merritt.cdlib.org>.

³³ This includes artifacts from the 2018-2019 LSTA-funded NAFAN planning initiative. These artifacts reflect the cumulative community-driven, research-based, exploration and development nature of this project.

1. **Research activity:** 1) iterative use of the research findings in the project's technical assessment and prototyping activity area; 2) publicly shared research findings for each prioritized area of investigation; 3) and feedback from the larger scholarly communication community based on conference presentations, responses to published works (which includes citations of research outputs).
2. **Technical assessment and prototyping activity:** 1) summaries of evaluations of relevant systems; 2) synthesized stakeholder feedback about the prototypes; and 3) feedback from the larger scholarly communications community about the systems evaluations, the prototype assessments, and the requirements for a minimum viable instantiation of the network. The research and technical assessment and prototyping activity areas are cross-informative, providing an evaluation mechanism to test understandings, implement prototypes, and adapt strategies as needed, in line with a pilot phase project.
3. **Community engagement activity:** 1) the extent of participation and resultant outcomes, and 2) the diversity of representation. As appropriate for a pilot phase, this indicator will allow evaluation and adjustment of the initial project structures where the Project Team and/or community stakeholders surface barriers to meaningful participation
4. **Sustainability and governance activity:** 1) market research into community capacity to support national network; 2) stakeholder responses to proposed sustainability and governance models; 3) formal agreements/commitments from aggregators to contribute time and/or money in the next phase; 4) degree of consensus on requirements by stakeholder groups; and 5) the level of continued interest in working/advisory group participation and willingness to take on governance responsibilities/opportunities. The Project Team will use these indicators to evaluate whether the project needs to adapt the timeframe and overall goals and ambitions.

Diversity Plan

As part of our process for recruiting researchers and archivists to participate in interviews and focus groups (see **Research data collection and analysis**), we will intentionally identify individuals at institutions that are underrepresented within existing aggregations. We will also identify archivists at institutions which predominantly focus on documenting the experiences of underrepresented or marginalized populations. These archivists can represent the needs of the communities and researchers they serve, as well as important considerations for describing and aggregating the collections they steward.³⁴

We will also create an open application process for participation in the Stakeholder Advisory Group, encouraging nominations and self-nominations from individuals who identify with underrepresented or marginalized populations as well as those whose work relates to underrepresented or marginalized populations. In developing the process, we will establish participation and recruitment criteria that ensure a diverse and representative range of perspectives and voices. In addition to sharing this invitation with focus group participants and distributing through aggregator communication channels, we will undertake targeted outreach to specific sections within the Society of American Archivists (Archives and Archivists of Color, Diverse Sexuality and Gender, Latin American and Caribbean Cultural Heritage Archives, and Native American Archives). We will also circulate the invitation to library, archive, and museum association mailing lists.

³⁴ The network will support policies and processes to promote inclusive descriptive practices, including support for multilingual and community-generated descriptions. For an example of recommendations for supporting community-based descriptive practices, see Black Lives in Philadelphia's Anti-Racist Description Working Group (2019), "Anti-Racist Description Resources," https://archivesforblacklives.files.wordpress.com/2019/10/ardr_final.pdf. For an example of initiatives that support multilingual approaches to archival description, see Sangawand, T-Kay (2020), "Ethical Archival Access in Cuba," https://stacks.stanford.edu/file/druid:wp525cy3737/2020_02_10_ltw_sangawand.pdf.

National Impact

Our pilot research and demonstration project will lay the foundation for meeting the critical equity and sustainability challenges facing researchers and cultural heritage institutions. A national finding aid network that is a community-driven, -sustained, and -governed will have a transformative and lasting impact by reducing barriers to discovery, expanding the historical record to include a broader range of voices, and establishing a solid, yet agile mechanism for solving shared infrastructure challenges going forward. In this vision:

- **Researchers** will efficiently discover and access a vastly larger set of holdings from an ever-increasing range of cultural heritage institutions documenting a much more representative set of communities, voices, and topics. These holdings will be integrated with related contextual information and digital content in complementary and interrelated national platforms (e.g., SNAC, DPLA).
- **Cultural heritage institutions** will participate in substantially greater numbers through supportive and equitable onramps for contributing finding aids, resulting in a more inclusive and representative platform for sharing information about collections. Their finding aids will gain wide exposure and be enhanced by proximity to related holdings from other institutions throughout the US. Those numerous institutions without a finding aid publication platform will benefit from a robust, hosted solution. All institutions, even those without the immediate capacity to create finding aids, will gain visibility by being listed, upon request, in an extensive, continually updated, and searchable registry of repositories that summarizes collecting areas and holdings information.
- **Existing statewide/regional finding aid aggregators** will invest in and utilize shared infrastructure, redirecting limited resources from costly, duplicative, systems into select services that bring the greatest value to their communities. They will engage with a greater number and range of cultural heritage organizations, supporting finding aid creation, addressing topical gaps and biases within the national aggregation, and enhancing legacy and existing finding aids to enable richer functionality.

Sustainability is a fundamental component of our project design, with research into viable sustainability and governance models prioritized to ensure a future beyond this initial funding period. Assuming positive outcomes in this pilot project, we intend to enact sustainable fiscal stewardship and governance strategies (likely including a combination of community support and grant funding), and scale our activities -- moving from a project to program -- in a subsequent development and implementation phase (2022-2024).

The project deliverables will be designed, wherever possible, with an eye toward re-use in other contexts related to developing and sustaining community-based scholarly communications infrastructure. The research findings will help establish a shared, updated understanding of user and contributor needs for archival discovery and access. The methodologies, data gathering instruments, and associated documentation will be managed in open source repositories for re-use. Similarly, prototype tools or demonstration systems developed during the course of the project will be managed in open source code repositories. We will also share in open access repositories, wikis or similar platforms, summaries of potential sustainability and governance model options, and associated benefits and risks, along with sample policies and operating principles.

The potential value of a fully-realized national finding aid network far exceeds the foundational IMLS funding investment requested in this proposal. This network represents an opportunity to leverage the original vision for finding aid aggregation into something bigger and more transformative. In this new period of growth and maturation, we will cultivate contributions from a broad array of aggregators and contributing institutions in order to provide the requisite financial and staffing resources to build and sustain the network for the long-term. We are committed to ensuring that previous investments by funding agencies in launching statewide/regional aggregations can be leveraged towards a shared and more sustainable solution, with a broad and lasting impact. Assuming positive indicators from this effort, CDL will continue to work with core partners to secure funding and resources for the next phase of work and will engage with the larger community to begin implementing the network and assembling the organizational structure to support it.

Schedule of Completion (Year 1)												
	Sep 20	Oct 20	Nov 20	Dec 20	Jan 21	Feb 21	Mar 21	Apr 21	May 21	Jun 21	Jul 21	Aug 21
Project management												
Establish Project Team and all Working and Advisory Groups												
Distribute subawards and sign consultant contracts												
Develop and launch project wiki and communication channels												
Hold Project Team online meetings; ensure early and continual findings sharing across Research and Technology assessment activity areas												
Research												
Hire research analyst												
Secure IRB approvals												
Hold Research Agenda Strategy Working Group online meetings												
Design, instrument, and deploy pop-up survey												
Design and conduct focus groups and interviews; share provisional findings with Project Team												
Transcribe, code and analyse focus group and interview data												
Conduct initial finding aid analysis; prepare report												
Report on initial findings at conferences (e.g., ALA/RBMS, CNI, SAA)												
Technical assessment and prototyping												
Hire developer												
Set up code repository												
Hold Technical Working and Advisory Group online meetings												
Support pop-up survey deployment; develop and implement processes to gather finding aid data												
Evaluate existing finding aid indexing/display systems and repository registry systems; share provisional findings with Project Team												
Develop prototype systems to demonstrate potential functional aspects of the network to Stakeholder and Working Groups and to support Research activities												
Community engagement												
Formulate and launch a communication/outreach strategy and develop supporting collateral (e.g., slide decks, talking points, etc.)												
Hold Aggregator Partners Working Group online meetings												
Hold Stakeholder Advisory Group online meetings												
Organize presentations at national and regional library, archive, and												
Sustainability and governance												
Segment stakeholder groups and markets												
Estimate operating cost scenarios with Project Team												
Define, develop, and conduct market research for each segment												
Evaluate market research; share provisional findings regarding income, participation, and governance models with Project Team; test with stakeholder groups												
Evaluate potential income models (collective, market transactions, hybrids); share findings with Project Team												

Schedule of Completion (Year 2)

	Sep 21	Oct 21	Nov 21	Dec 21	Jan 22	Feb 22	Mar 22	Apr 22	May 22	Jun 22	Jul 22	Aug 22
Project management												
Hold Project Team online meetings												
Research												
Hold Research Agenda Strategy and Working Group online meetings												
Transcribe, code and analyse focus group and interview data												
Document and publish interview and focus group findings												
Conduct second phase finding aid data analysis; prepare report												
Report on findings at conferences (e.g., ALA/RBMS, CNI, SAA)												
Technical assessment and prototyping												
Hold Technical Working and Advisory Group online meetings												
Prepare public report of results from assessment of existing finding aid indexing/display systems and repository registry systems												
Develop prototype systems (as needed) to demonstrate potential functional aspects of the network to Stakeholder and Working Groups and to support Research activities												
Formulate requirements, timeline, and plan for minimum viable instantiation of the network (with Research team), post 2022												
Community engagement												
Hold Aggregator Partners Working Group online meetings												
Hold Stakeholder Advisory Group online meetings												
Organize presentations at national and regional library, archive, and museum association conferences												
Sustainability and governance												
Evaluate potential income models (collective, market transactions, hybrids); share findings with Project Team												
Propose governance and organizational structures appropriate to income model												
Test and refine financial projections												
Identify a potential network convener/organizational home (or homes) for 2022-2024 development phase												
Secure formal agreements/commitments from aggregators to support 2022-2024 implementation phase												



DIGITAL PRODUCT FORM

INTRODUCTION

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to digital products that are created using federal funds. This includes (1) digitized and born-digital content, resources, or assets; (2) software; and (3) research data (see below for more specific examples). Excluded are preliminary analyses, drafts of papers, plans for future research, peer-review assessments, and communications with colleagues.

The digital products you create with IMLS funding require effective stewardship to protect and enhance their value, and they should be freely and readily available for use and reuse by libraries, archives, museums, and the public. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

INSTRUCTIONS

If you propose to create digital products in the course of your IMLS-funded project, you must first provide answers to the questions in **SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS**. Then consider which of the following types of digital products you will create in your project, and complete each section of the form that is applicable.

SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS

Complete this section if your project will create digital content, resources, or assets. These include both digitized and born-digital products created by individuals, project teams, or through community gatherings during your project. Examples include, but are not limited to, still images, audio files, moving images, microfilm, object inventories, object catalogs, artworks, books, posters, curricula, field books, maps, notebooks, scientific labels, metadata schema, charts, tables, drawings, workflows, and teacher toolkits. Your project may involve making these materials available through public or access-controlled websites, kiosks, or live or recorded programs.

SECTION III: SOFTWARE

Complete this section if your project will create software, including any source code, algorithms, applications, and digital tools plus the accompanying documentation created by you during your project.

SECTION IV: RESEARCH DATA

Complete this section if your project will create research data, including recorded factual information and supporting documentation, commonly accepted as relevant to validating research findings and to supporting scholarly publications.

SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS

A.1 We expect applicants seeking federal funds for developing or creating digital products to release these files under open-source licenses to maximize access and promote reuse. What will be the intellectual property status of the digital products (i.e., digital content, resources, or assets; software; research data) you intend to create? What ownership rights will your organization assert over the files you intend to create, and what conditions will you impose on their access and use? Who will hold the copyright(s)? Explain and justify your licensing selections. Identify and explain the license under which you will release the files (e.g., a non-restrictive license such as BSD, GNU, MIT, Creative Commons licenses; RightsStatements.org statements). Explain and justify any prohibitive terms or conditions of use or access, and detail how you will notify potential users about relevant terms and conditions.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and the format(s) you will use.

A.2 List the equipment, software, and supplies that you will use to create the digital content, resources, or assets, or the name of the service provider that will perform the work.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG, OBJ, DOC, PDF) you plan to use. If digitizing content, describe the quality standards (e.g., resolution, sampling rate, pixel dimensions) you will use for the files you will create.

Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan. How will you monitor and evaluate your workflow and products?

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period. Your plan should address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata or linked data. Specify which standards or data models you will use for the metadata structure (e.g., RDF, BIBFRAME, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content, delivery enabled by IIIF specifications).

D.2. Provide the name(s) and URL(s) (Universal Resource Locator), DOI (Digital Object Identifier), or other persistent identifier for any examples of previous digital content, resources, or assets your organization has created.

SECTION III: SOFTWARE

General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

A.2 List other existing software that wholly or partially performs the same or similar functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

Technical Information

B.1 List the programming languages, platforms, frameworks, software, or other applications you will use to create your software and explain why you chose them.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

B.5 Provide the name(s), URL(s), and/or code repository locations for examples of any previous software your organization has created.

Access and Use

C.1 Describe how you will make the software and source code available to the public and/or its intended users.

C.2 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

URL:

SECTION IV: RESEARCH DATA

As part of the federal government's commitment to increase access to federally funded research data, Section IV represents the Data Management Plan (DMP) for research proposals and should reflect data management, dissemination, and preservation best practices in the applicant's area of research appropriate to the data that the project will generate.

A.1 Identify the type(s) of data you plan to collect or generate, and the purpose or intended use(s) to which you expect them to be put. Describe the method(s) you will use, the proposed scope and scale, and the approximate dates or intervals at which you will collect or generate data.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

A.3 Will you collect any sensitive information? This may include personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information. If so, detail the specific steps you will take to protect the information while you prepare it for public release (e.g., anonymizing individual identifiers, data aggregation). If the data will not be released publicly, explain why the data cannot be shared due to the protection of privacy, confidentiality, security, intellectual property, and other rights or requirements.

A.4 What technical (hardware and/or software) requirements or dependencies would be necessary for understanding retrieving, displaying, processing, or otherwise reusing the data?

A.5 What documentation (e.g., consent agreements, data documentation, codebooks, metadata, and analytical and procedural information) will you capture or create along with the data? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the data it describes to enable future reuse?

A.6 What is your plan for managing, disseminating, and preserving data after the completion of the award-funded project?

A.7 Identify where you will deposit the data:

Name of repository:

URL:

A.8 When and how frequently will you review this data management plan? How will the implementation be monitored?