Toward a National Archival Finding Aid Network

From Planning Initiative to Project and Program: An Action Plan

October 2019

Action Plan Task Force
Jodi Allison-Bunnell, AB Consulting
Bradley Daigle, Virginia Heritage
Elizabeth Dunham, Arizona Archives Online
Karen Eberhart, Rhode Island Archives and Manuscripts Online
Brian Tingle, Online Archive of California
Adrian Turner, Online Archive of California (Chair)

Table of Contents

Executive Summary and Recommendations

After two decades of operation, finding aid (or collection description) aggregations across the U.S. are now struggling to find sufficient resources to update aging infrastructure, meet evolving user needs, and engage with some of the most promising advances in the field. In 2018-2019, the California Digital Library (CDL) organized an LSTA-funded planning initiative, "Toward a National Archival Finding Aid Network (NAFAN)," convening U.S. finding aid aggregators and expert advisors with deep knowledge of organizational development, community engagement, and sustainability.[1] Preparatory research and outcomes from a planning symposium clearly demonstrate that aggregations are at-risk in their current siloed contexts.[2] Participants broadly affirmed that it is time to rethink aggregation at scale, calling for a substantially more inclusive framework that enables participation by a wider range of cultural heritage institutional contributors, supports a greater variety of collection description levels and formats, and meets the needs and interests of a more diverse set of end users--while simultaneously transitioning away from outmoded technologies and directly addressing foundational issues of sustainability.

This action plan is a key deliverable of the NAFAN planning initiative, drawing directly on the symposium outcomes. The plan was prepared by a Task Force comprising representatives from the Core Partner group of aggregators, who contributed time between July-September 2019 to formulate and develop these recommendations.

At the heart of the action plan are recommendations for and principles to guide next steps to implement a national-level finding aid network. The Task Force recommends a phased, incremental approach that moves this effort from a research and demonstration project to a program; is informed by a research agenda; and (from the beginning) includes work to establish business and governance models that fit the infrastructure and service model. At a high level, the phases and their goals/purpose are as follows:

- **Immediate (2019 Fall-2020 Summer):** Sustain the effort that began with the 2018-2019 planning initiative in order to enable the near- and longer-term phases, including sustained community engagement, articulation of the research agenda and high-level requirements for the project, and targeted grant applications.
- **Near-term (2020 Summer-2022):** Demonstrate value of scaling current siloed aggregation activities to a national level, and implement a research agenda to guide the mid- to long-term development of the network. Generate engagement through activities that establish a vision for the network, produce near-term value, and lay the foundation for mid- and long-term value.
- **Mid-term (2022-2024):** Transition from project to program, using findings from the research agenda. Move from demonstration mode to building mode, in a solidified partner/community structure.
- **Long-term (2024 and beyond):** Realize a fully established network that supports a broad spectrum of contributors and provides transformative access for end-users.

Each phase includes action items related to different aspects of growing the network: community building, iterative technical and services development, research agenda, business and sustainability

---

[1] Toward a National Archival Finding Aid Network project wiki: https://confluence.ucop.edu/display/NAFAN/

[2] See Allison-Bunnell, J. (2019). "Finding Aid Aggregation at a Crossroads," https://escholarship.org/uc/item/5sp13112 . See also California Digital Library. "Toward a National Archival Finding Aid Network: Summary of Outcomes from Symposium," https://docs.google.com/document/d/1sd-q59u_L-NHhZG1Ib8hCFBsg2BG0a0WuwNwsln5TSo/edit?usp=sharing

models, coordination and governance, and resourcing. For further details on all of these, please see the **Action Plan Phases** section of this document.

The action plan also includes a summary of high-level requirements and functions envisioned for the network, and a summary of questions to explore as part of the research agenda. See **Appendix A: Proposed High-Level Requirements and Functions** and **Appendix B: Research Agenda** for more information.

Action Plan Phases

	Immediate (2019 Fall-2020 Summer)	Near-term (2020 Summer-2022)	Mid-term (2022-2024)	Long-term (2024 and beyond)
Goal / Purpose	Sustain the effort that began with the 2018-2019 planning initiative in order to enable the near- and longer-term phases, including sustained community engagement, articulation of the research agenda and high level requirements for the project, and targeted grant applications.	Demonstrate value of scaling current siloed aggregation activities to a national level, and implement a research agenda to guide the mid- to long-term development of the network. Generate engagement through activities that establish a vision for the network, produce near-term value, and lay the foundation for mid- and long-term value.	Transition from project to program, using findings from the research agenda. Move from demonstration mode to building mode, in a solidified partner/community structure.	Realize a fully established network that supports a broad spectrum of contributors and provides transformative access for end-users.
Community Building	● Keep the community of aggregators engaged and invested in advancing the near- to long-term activities. ● Keep the broader archives community informed using coordinated talking points/messaging, presentations at key conferences, etc.	● Formalize communication mechanisms across the community of interest, ensuring stable mechanisms to help entities work together to address shared areas of need. ● Establish and implement a communication/outreach strategy to build broad awareness about our work; share information on timelines and development phases; manage expectations. ● Promote the value proposition beyond the aggregator and archival community to other related areas and activities (e.g., scholarly communications, open access publishing initiatives).	● Strengthen and expand participation by bringing on more aggregators and (as appropriate) other entities. ● Develop transition models for aggregators. ● Evaluate community engagement levels; address gaps.	● Continue to expand participation by aggregators and institutions. ● Continue to refine community engagement.

	● Track and facilitate sharing about transition planning for existing statewide/regional platforms.		
Iterative Technical and Services Development	● Scope prototyping activities to : 1) assess the feasibility of using different technologies and approaches to build components/parts of the network, and 2) facilitate discussion and evaluate work regarding targeted requirements . (See **Appendix A: Proposed High-Level Requirements and Functions**).	● Conduct due diligence evaluation of existing systems, including scholarly publishing platforms, to see if any meet requirements for a minimum viable product. ● Drawing on initial research agenda findings, prototype components/parts of the network. Assess the feasibility of using different technologies and approaches.. (See **Appendix A: Proposed High-Level Requirements and Functions**). ● Generate engagement through prototyping activities that demonstrate vision, as well as near-, mid-, and long-term value. ● Explore branding strategies for the network.	● Design and develop a minimum viable product, comprising a cost-efficient, durable, and robust technical solution informed by research agenda findings. (See **Appendix A: Proposed High-Level Requirements and Functions** and **Appendix B: Research Agenda**). ● Continue ongoing prototyping, development, and evaluation of additional technical features. ● Build out the functionality and services provided via the network's social/organizational infrastructure (e.g. training, outreach, support). ● Define shared service models (where some services may be centralized vs. distributed). ● Develop migration paths from existing statewide/regional platforms to the network.	● Establish an ongoing development plan for the network, with assessments and regular feature/enhancement releases. ● Support broader adoption of shared infrastructure and service models. ● Increase scale to support a broad range and large number of contributors.

Research Agenda	● Scope and refine research agenda. (See **Appendix B: Research Agenda**).	● Support/launch the research agenda. (See **Appendix B: Research Agenda**). ● Identify additional entities within the community to support or participate in research agenda. ● Fold research findings into relevant activities.	● Collectively ensure progress addressing the research agenda, along with sharing and discussion of findings. (See **Appendix B: Research Agenda**). ● Develop strategies for identifying and addressing ongoing research needs.	● Maintain the relevancy of the network, through ongoing research and engagement with contributor and end-user stakeholders.
Business and Sustainability Models	● Determine technical resources needed for near-term phase activities. ● Determine scope for business/market analysis. ● Test value proposition of a fully realized network against hypothetical operating cost requirements.	● Determine costs for building and sustaining a minimum viable product. ● Conduct business/market analysis to surface sources of resources for startup, ongoing costs, and development; identify most promising options. ● Investigate resourcing opportunities/funding lines beyond standing archives/special collections budgets. ● Conduct gap analysis between aspirations and the available resources stakeholders are able to commit; adjust or phase plans accordingly. ● Establish ongoing practices for creating and revising success measures, assessing value propositions and resource capacities among key stakeholders.	● Implement agreed-upon business models. ● Establish on-going practices for assessing value propositions and resource capacities among key stakeholders.	● Monitor and refine business models. ● Continue to assess value proposition, success measures, and process for responsibility and revision.

Coordination and Governance	● Establish a formal partnership of organizations that will participate in co-leadership, research, and/or development work on near-term activities. ● Determine interim convener/organizational home(s). ● Establish an interim advisory group to facilitate and maintain community engagement, and advise on development of the network. ● Specify modes and opportunities for the community of aggregators to support and/or participate in near-term activities.	● Build requirements and criteria for long-term organizational home(s); identify and assess options. ● Formulate governance models and initial policies.	● Establish long-term organizational home(s). ● Establish appropriate governance models and policies.	● Monitor and refine business models. ● Monitor and refine governance models and policies.
Resourcing	● Prepare one or more grant proposals to facilitate work on near-term activities.[3] ● Explore bridge funding opportunities to support immediate planning efforts. ● Confirm in-kind contributions from an initial group of research and development partners. ● Coordinate funding requests and projects among partners.	● Kickstart with grant / seed funding and in-kind contributions from an initial group of research and development partners.	● Support activities through grant(s), in-kind contributions, and early implementation of the business models.	● Ensure sufficient resourcing through the adoption of and adherence to appropriate business and governance models.

---

[3] The Task Force also prepared the following additional working document identifying potential sources: "Funding opportunities to support aspects of near- and mid-term phases," available at https://confluence.ucop.edu/display/NAFAN/Reports+and+Reference+Resources

Principles to Guide Next Steps

The following principles were derived from the outcomes of the symposium and will guide further collaboration:

- We will base our long term actions on our research agenda rather than current assumptions.
- We will scale discovery and access to a national level that is as comprehensive as possible.
- We will increase usable access to unique resources for the broadest possible range of users.
- We will support the broadest possible range of contributing institutions and minimize barriers to participation.
- We understand that a sustainable collaboration is one that is supported with time, expertise, and resources, and that we will need a range of stakeholders -- including aggregators and individual contributing institutions -- to realize the long-term vision.
- We understand that perfection is the enemy of the good and will abandon it in service of focused and logical investment.

Appendix A: Proposed High-Level Requirements and Functions

High-Level Requirements

Among the outcomes of the symposium are the following requirements for the long-term vision for a national-level finding aid network. At this stage of development, and with a significant research agenda to be enacted, technical or infrastructure requirements would be premature. Instead, the points below address the value propositions, broader purpose and outcomes, and provide the underpinnings of a sustainable enterprise.

Serve End Users
- Facilitate direct and ongoing involvement of a diverse range of researchers end users in order to shape creation and development of the system.
- Support discovery that gets end users as close to the full resource as possible.
  - The ideal is the actual resource, or digital representation thereof, in context.
  - Provide maximum integration between collection descriptions (with related context) and content.

Be Inclusive and Comprehensive
- Recognize the value of under-represented contributors and collections and create low barriers to entry.
- Flexibly support the participation of contributors who cannot make their holdings available online, due to legal, ethical, cultural, and other factors.

Reduce Local Work for Contributors
- Support low barriers to entry for contributors.
  - Take many forms of existing metadata (e.g., structured in formats such as EAD and MARC, unstructured full-text within PDF files)
  - Set minimal requirements for descriptive metadata (e.g., based on *Describing Archives: a Content Standard's* single-level minimum requirements[4]) to sufficiently support browsing/searching.
  - Integrate with related tools (e.g., ArchivesSpace, Access to Memory) to support efficient workflows.
  - Automate the contribution process so that it is as easy as possible.
  - Allow for iteration over time.
- Clearly identify the contributing institution.
- Recognize institutional investment through branding that supports good user experience.

---

[4] Society of American Archivists (2013). "Describing Archives: a Content Standard" (DACS), https://www2.archivists.org/standards/DACS

High-Level Functions for a Fully-Realized Network

Building on discussions from the symposium, the Task Force proposes the following high-level functions for a fully-realized national-level finding aid network, acknowledging that there will likely be changes over time. A subset of these functions would be demonstrated in the near and mid-term prototyping and minimum viable product development phase, and the full suite of functions would be represented within the long-term development phase:[5]
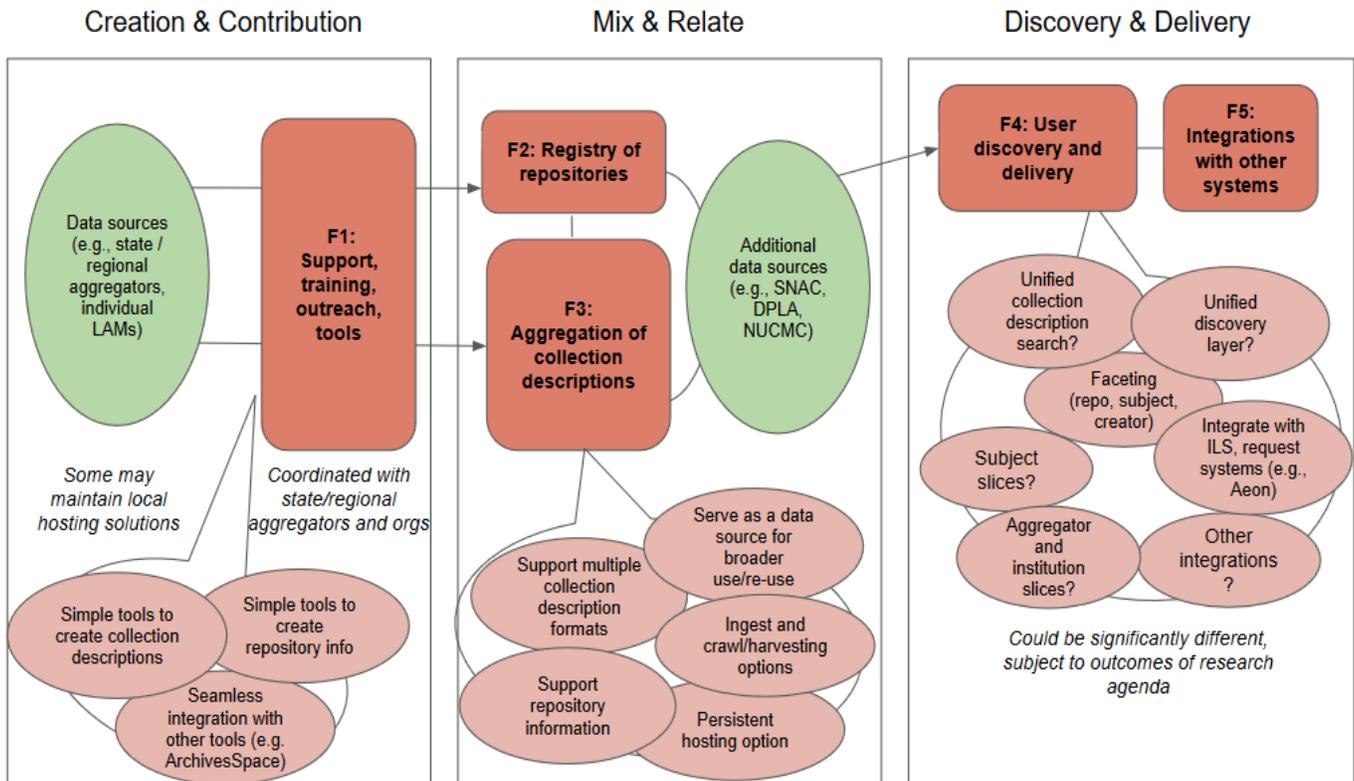
● = Example function   **● = Example requirement based on the function**   **● = Example contributor and data source**

Function 1: Support, training, outreach, and tools

Provide tools, regional support, outreach, and training, to enable a broad range of repositories to expose their collection guides through the network. Geographically distributed regional support could potentially be provided by a range of entities, including state and local collaborative networks, DPLA Hubs, state libraries, State Historic Records Advisory Boards (SHRABs), and state/regional library, archives, and museum associations.

---

[5] The Task Force also prepared the following additional working documents, establishing baseline requirements for a shared national-level infrastructure (to serve as a viable alternative to existing statewide/regional platforms): "Sample user scenarios and personae," and "Summary of core functions represented by statewide/regional aggregator platforms," available at https://confluence.ucop.edu/display/NAFAN/Reports+and+Reference+Resources

Function 2: Registry of repositories

Allow any repository to be listed in a national directory, whether they contribute collection descriptions or not.

Example requirements
● Persistent management of standardized repository information (e.g., based elements in the *International Standard for Describing Institutions with Archival Holdings*[6]). Includes address, contact info, and geographic coordinates. Hours and request forms if available. Also may include collecting areas, policies, etc.
● Provides unique identifiers and authority control for repositories.
● Contains a machine readable description of how the platform is supposed to process and index repository records, ingested and/or crawled/harvested from external systems.

Function 3: Aggregation of collection descriptions

Provide a robust, comprehensively aggregated, publishing platform for collection descriptions.

Example requirements
● Provide shared infrastructure and service models, to support aggregation of collection descriptions at a lower cost than existing regional/statewide models.
● Support ingest of collection description in a range of different file formats, e.g., Encoded Archival Description (EAD), MARC, PDF (potentially also other formats, e.g., Word, Excel).
● Support crawling/harvesting of collection descriptions in a range of formats, published in external systems.
● Support seamless ingest of collection descriptions, contributed from external systems (e.g., ArchivesSpace)
● Support validation and/or minimal requirements checking.
● Support persistent hosting of collection descriptions.
● Provide persistent identifiers for each collection description.
● Provide options to expose or re-use data within the central index, within other contexts.

Function 4: User discovery and delivery

Support discovery, use, and re-use of collection descriptions, based on prioritized user needs identified through research activities.

Example requirements
● Enable discovery through commonly used search engines.
● Provide a unified index of collection descriptions, whether ingested or harvested/crawled.
● Support full text and fielded search of collection descriptions.
● Support faceting by repository, subject, and collection creators/contributors.
● Facilitate in-person use of analog materials at the contributing institution.
● Provide persistent and accurate repository information.

---

[6] International Council on Archives (ICA): "ISDIAH: International Standard for Describing Institutions with Archival Holdings" (2011): https://www.ica.org/en/isdiah-international-standard-describing-institutions-archival-holdings

Function 5: Integrations with other systems
Support integration of collection descriptions with context and content in other systems, based on prioritized user needs identified through research activities.

Example requirements
- Deliver integrated context and content from many sources to provide access that aligns with user needs.
- Provide linkages to and/or integrations with digital content from archival materials that matches the user's search.
- Show related context from Social Networks and Archival Context (SNAC): records for corporations, persons, and families.
- Provide options to expose or re-use data within Integrated Library Systems (ILS) and discovery systems.
- Provide linkages to and/or integrations with reference or request systems such as Aeon.

Appendix B: Research Agenda

In both the preparatory research and symposium outcomes, we identified gaps in current knowledge that must be addressed in order to create a user-centered, high-value, sustainable approach to facilitating access to archival description, content, and context. With a stated commitment to pursue a research agenda and to ensure that research findings are the main driver for shared decision-making, research is both critical to and the first priority for further collective action.

Scholar and Other End-User Stakeholder Needs

One of the high-level findings from the symposium summary is the concerning fact that, despite twenty years of effort, users still cannot locate what they need. User needs assessment will be foundational to developing our long-term vision. It will provide a vital framework for envisioning infrastructure, communications, community building, and funding requests. What we know for certain is that access is currently based on arbitrary state, regional, and format silos that are not comprehensive. Moreover, access is not intuitive -- and collection descriptions lack integration with highly related context and content. Hence, we have identified the following value that aggregation should provide to users:

- Help researchers find what they need by providing more "about"-based access to collections (e.g., based on topic, context pertaining to the creators or contributors to the collection, etc.).
- Provide integration of collection descriptions with related digital content and context.

Thus, the highest priority research questions are those that will firmly identify high-value user-based functions and requirements, to be implemented in development cycles over the entire course of the action plan. This systematic study of end users will pursue the following questions:

- By aggregating collection descriptions, do we enable broader access to and usage of collections by researchers?
- What do we already know about the value of providing access to collections via collection descriptions? Conduct a thorough review of the existing literature.
- Who are our researchers? We need to understand the range of users, particularly beyond academic users, and their motivations.
- How are they looking for things? How are they trying to access our collections? How would they like to use, share, and re-use collection information?
- Based on all available analytics, how are users utilizing collection descriptions within current aggregations? Does aggregation provide a valuable service to end users? How does it compare with the use of digital content and context?
- How does the potential integration of archival description, content, and context support user needs? How do we best and most efficiently integrate collection descriptions with related context and content in external systems?
- Which system integrations provide sufficient value to undertake the work needed to support those needs?
- How much description and data structure is required to support the outcomes we determine that end users need?
- What automation strategies can be leveraged to enable this?

From their earliest stages, these user studies should inform the iterative development of system features. They should also facilitate innovation and exploratory approaches that may be radically different from current finding aid presentations.

Additionally, based on these user studies, we will articulate the societal value of making our collections available for use -- and the specific role of aggregation in creating that value. Our hypothesis is that aggregation promotes equity and makes it easier to discover collections, but we will test this theory with more specific and actionable information.

Contributor Participation and Needs
The preparatory research and symposium outcomes suggest that aggregators currently provide valued services to contributing institutions, by providing additional exposure for their collections, offering a persistent hosting solution, and cultivating a community of practice.

However, we know that preparing and contributing collection descriptions—particularly EAD-formatted finding aids—is too high a bar for many institutions; because of resource limitations or institutional preferences, they often do not and will not create them. In cases where institutions can generate EAD files, the process of preparing, submitting, and maintaining the files can be burdensome. We thus identified our goals for supporting contributing institutions as:

- Provide low-barrier entry to aggregation that works with a variety of collection description types.
- Reduce local work.
- Build community within and across aggregators.
- Re-use existing collection description metadata.

Thus, the highest priority research questions should identify barriers to and incentives for participation, in order to expand the range of contributors and collections that could be supported by a national-level finding aid network:

- What are the key barriers institutions face when describing their collections? When contributing to aggregators?
- What formats should be supported to engage the greatest number of contributing institutions efficiently?
- What are the minimal levels of collection/item description (beyond *Describing Archives: a Content Standard's* single-level minimum requirements) that will ensure discovery, provide a positive end user experience, and keep barriers to contributing institutions low? What is the minimal level/quality of description that still supports leveraging the latent power and promise of aggregation?[7]
- What is the current extent of data structure and content consistency, across extant collection descriptions?
- How can we support contributing institutions in achieving the necessary level of description?
- What tools and services might we provide to support the creation and sharing of simple collection descriptions (including simple container lists/inventories) and repository information?
- Within the context of archival collection management systems such as ArchivesSpace: what kinds of outputs can be generated? Should we consider ingesting and/or crawling/harvesting other kinds of outputs beyond finding aid data (e.g., accession records)?
- What are the levels and type(s) of institutional signaling (attribution, branding) that bring the most value to stakeholders?

---

[7] We also anticipate scenarios where different levels of collection description contribution could correlate to different levels of display services and features, based on the level and quality of source data.

- What existing sources could be leveraged to supply information for a comprehensive registry of repositories? What do those sources have or lack?
- What existing robust systems could serve as a platform for indexing and displaying finding aids? Assess both aggregator systems that are considered robust as well as other types of publication platforms to see which offer the most promising options.
- What is the most viable model for aggregating collection descriptions? Is it essential to support only contribution or ingest to a central infrastructure, or is there value in also supporting a crawl or harvest model that allows for local hosting and presentation at the institution?
- What form or forms of persistent identifiers do aggregators and institutions currently use? What do we need persistent identifiers to do? Which of those would be appropriate for a national-level finding aid network?