

## Building Archival Digital Discovery and Access Systems with Arclight

**Introduction:** The University at Albany, SUNY (UAlbany) and the Empire State Library Network (ESLN) request \$249,999 for a two-year implementation grant to further develop Arclight into an access system for digital archives and special collections materials. This project will enable Arclight to provide discovery and access to digital materials, metadata, and full-text content using International Image Interoperability Framework (IIIF) manifests, and lowers the barriers for under-resourced institutions to implement systems for archival discovery and delivery. This work is well-aligned with the NLG-L Program Goal 3 and also builds on the knowledge and accomplishments of three IMLS-funded projects, the [Cross-Search and Context research project](#), the [Lighting the Way National Forum](#), and the [National Finding Aid Network](#) project, which found that users struggled to access digital content described by finding aids.

**Project Justification:** Archivists use hierarchical finding aids to manage large volumes of material at progressive levels of detail based on value and use. However, archivists currently struggle to provide access to digital objects using the same methods. Current digital repositories or digital asset management systems are difficult to integrate with archival systems and often require an additional discovery point for users. ESLN hosts two statewide systems for over 400 institutions: [EmpireADC](#) for archival finding aids and [New York Heritage](#) for digital objects. Contributors, who are often small and under-resourced, must separate materials or even duplicate their archival description to display in both systems. In practice, digital repositories require additional metadata records in a separate system, which creates unnecessary labor for digitized items and often makes detailed access to born-digital records impossible. All institutions struggle to integrate digital repositories with archival finding aids, requiring complex workflows to span multiple systems. Users must then navigate multiple systems, often for items that might be right next to each other in a physical or digital folder.

[Arclight](#) is currently a discovery system for archival finding aids, but since it is based on [Blacklight](#), it can also be used to provide discovery and access for digital objects, including full-text OCR and transcription content. Since institutions use a wide variety of digital repository systems and metadata schemas, the biggest barriers to including digital objects in Arclight are the lack of common data pipelines and harvesting tools. Using IIIF manifests to make these connections will ensure that Arclight can work with a wide variety of existing digital repositories and other emerging solutions. Since most [current Arclight adopters](#) are major research libraries, this project aims to lower implementation and maintenance barriers for small institutions that struggle to adopt *yet another* system. Joining finding aids, digital object metadata, and full text content in a single discovery system will pose usability challenges as well as create new opportunities. To address this, the project will experiment with weighting and inheriting description in Arclight in new ways according to [archival theory and standards](#) and assess its effectiveness using iterative user experience testing.

**Project Work Plan:** The Project Team will include two archivists and a library technologist from UAlbany, the Project Coordinator for EmpireADC from ESLN, and representatives from the four pilot partners. The team will meet regularly with an Advisory Board of practitioners drawn from existing Arclight implementations at major research libraries. Deliverables include open specifications for modeling digital objects in Arclight, an ArchivesSpace plugin to index data directly to Arclight with digital object data from IIIF manifests, a command line harvesting tool for aggregators, implementations at EmpireADC and UAlbany, and a report of user experience testing results.

**Phase One:** The project requires a collaborative community effort to better define how digital objects fit conceptually within archival description, as this is currently undertheorized with substantial variations in practice. The project will solicit a cohort of practitioners with expertise working with archival data from a public call to the Code4Lib, DLF, and SAA announcements listservs. A larger group of 28 will meet remotely in October 2024 to participate in facilitated exercises, contribute examples and use cases, and outline initial principles and data structures. In December 2024, a subgroup of 8 will meet in-person for three days to draft both a generalized data model and a system-specific Arclight Solr Index Specification that defines how digital objects, full-text content, and digital object metadata will be managed by Arclight in practice. In February 2025, the full cohort will review the model and specification. In addition, the project team will create an Arclight IIIF Specification that will define specific requirements for exposing full-text content and digital object metadata in a manifest for harvesting into Arclight based on established practices in the IIIF community. These documents will be made available for public comment before the project team publishes final versions in May 2025.

**Phase Two:** In April-October 2025, UAlbany will develop two tools for harvesting data into Arclight that will parse linked IIIF manifests to incorporate digital object metadata and transcriptions in the Arclight index. A vendor will develop the ArchivesSpace plugin to index data directly into Arclight as archivists edit and save resources and archival objects. UAlbany will also build out the existing [description indexer tool](#), which will provide a command line utility for aggregators to harvest data from Encoded Archival Description, the ArchivesSpace API, and ArchivesSpace OAI-PMH feeds. UAlbany will make updates and provide public documentation for implementing a IIIF viewer in Arclight to display digital objects.

**Phase Three:** UAlbany and ESLN will implement Arclight as a single discovery system for both archival description and digital objects, both locally at UAlbany and for EmpireADC. The ESLN effort will focus on two small institutions, Hudson Area Library and Historic Huguenot Street that host digital objects in New York Heritage, as well as two medium-sized repositories, Rensselaer Polytechnic Institute and Union College that use Archipelago as a digital repository. These partners will harvest portions of their archival description and digital objects into EmpireADC as pilot projects. UAlbany's implementation will not use a traditional digital repository, but will manage digital objects on network file shares according to an open local specification. These objects will be served using a IIIF image server and harvested into Arclight.

**Phase Four:** In spring 2026, the project team will initiate an iterative user testing process for Arclight instances with both archival description and digital content. We have had preliminary conversations with the User Experience Section of SAA to assist and/or participate. We will design an open national call for contributing archivists to provide example description and end-user discovery and navigation tasks for testing. Examples and tasks with digital objects will also be developed by UAlbany and the four ESLN pilot partners. End-users will be solicited to perform tasks remotely over Zoom, and contributing archivists will both view the sessions and be able to test and provide comments on the test Arclight instances themselves. We will publish a report of our findings in an open access repository or journal.

**Diversity Plan:** The project team acknowledges that its white, cisgender backgrounds are heavily overrepresented in the archives profession, particularly among archivists that work with technology. In selecting cohort participants, the project will prioritize and make targeted outreach to members of marginalized and/or underrepresented communities to both ensure a diversity of voices and help professionals from a broader variety of backgrounds gain confidence and experience in this space. The project is committed to care-focused and anti-oppressive facilitation methods that attempt to make a welcoming space for contributors from diverse backgrounds and experiences.

**Project Results:** This project makes a national-level impact by enabling Arclight to be a single point of discovery for digital objects – allowing institutions to preserve both archival context and valuable descriptive labor. It also enables digitization on-demand as archivists can use existing description instead of creating new metadata, and allows access to born-digital records which require varying levels of description and access that is challenging with traditional digital repositories. As Arclight takes on the role of access to digital objects, this work will also lower the bar for digital repositories and create new possibilities. UAlbany's local implementation will manage digital objects and provide IIIF manifests from network shares according to a local specification, an approach that may be easier to maintain than traditional digital repositories. Currently, the technical and organizational costs of making digital materials available are so onerous that small archives—which are most likely to hold histories of under-represented groups—find themselves in a difficult position. They are often forced to use systems that were not built for archival materials, which means that they do not benefit from collective efforts and often fall behind when these systems fall into obsolescence. With this work, consortia like ESLN that are focused on under-resourced institutions can provide participants with a single point of access for archival description and digital objects. This work will also be incorporated into existing Arclight instances at major research libraries, and the ArchivesSpace plugin will make it easier for medium-sized repositories to implement Arclight. Thus, software improvements made by better-resourced institutions will be available to institutions of all sizes.

**Budget Summary:** The project requests \$249,999, including \$36,238 for salaries and \$23,972 fringe benefits (PI at 25% for two years), \$9,600 for travel (8 member in-person cohort), \$92,814 in subawards and contracts (\$70,000 for vendor development and \$22,814 subaward for an ESLN representative at 10% for two years), \$25,345 in other costs (\$24,500 in honoraria, \$1,000 each for 8 member in-person cohort, and \$500 each for 20 member remote cohort and 6 member advisory board, and \$845 in EmpireADC fees), and \$62,030 for 33% indirect costs.