

# Finding Aids Unleashed

## Iterative Development of a Portable Publication System

Deb Verhoff, Joseph G. Pawletko, Donald R. Mennerich, and Laura Henze

---

### ABSTRACT

New York University Libraries recently completed a redesign for their finding aids publishing service to replace an outdated XSLT stylesheet publishing method. The primary design goals focused on accessibility and usability for patrons, including improving the presentation of digital archival objects. In this article, we focus on the iterative process devised by a team of designers, developers, and archivists. We discuss our process for creating a data model to map Encoded Archival Description files exported from ArchivesSpace into JSON structured data for use with Hugo, an open-source static site generator. We present our overall systems design for the suite of microservices used to automate and scale this process. The new solution is available for other institutions to leverage for their finding aids.

### INTRODUCTION

New York University Libraries has long provided archivists with a service to support web publishing for finding aids, descriptions of archival collections in the Encoded Archival Description (EAD) schema. By 2020, the software behind this service had become unsustainable due to the reliance on a single XSLT stylesheet to publish over 5,000 unique finding aids for New York University and two partner institutions, the Center for Brooklyn History and The New York Historical. The eXtensible Stylesheet Language Transformation (XSLT) process was brittle: modifying the stylesheet for one archival repository caused problems with publishing finding aids for other repositories. Also, the resulting web-based finding aids generated by the XSLT process had usability problems for patrons and issues with the presentation of digital archival objects.

From 2020 through 2023, New York University Libraries conducted a project to replace the outdated publishing software and address the usability and digital archival object presentation issues. The project resulted in a new method for transforming EAD XML files into JSON formatted data, which can be used in multiple publication and discovery environments. The new solution uses modern open-source technologies and is available for other institutions to leverage for their finding aids.

A project team comprised of three developers, two archivists, and a designer was formed to address two primary objectives: (1) create a new publishing system for archivists to preview and publish finding aids to the web, and (2) improve the visual design for the published finding aids. Design objectives focused on improving accessibility and usability for patrons.

#### About the Authors

**Deb Verhoff** ([dv47@nyu.edu](mailto:dv47@nyu.edu)) (corresponding author) is Digital Collections Manager, New York University. **Joseph G. Pawletko** ([jgp@nyu.edu](mailto:jgp@nyu.edu)) is Senior Software Developer, New York University. **Donald R. Mennerich** ([dm3053@nyu.edu](mailto:dm3053@nyu.edu)) is Senior Digital Archivist, New York University. **Laura Henze** ([lh435@nyu.edu](mailto:lh435@nyu.edu)) is Senior UX Designer, New York University. © 2025.

Submitted: 19 December 2024. Accepted for Publication: 15 June 2025. Published: 15 September 2025.

The project partners represented nine distinct archival repositories where archival description practices varied. The daunting scale of thousands of finding aids initially created a sense that the variability was too vast to accommodate. For this reason, the project team embraced a method of working iteratively.

## BACKGROUND AND CONTEXT

In 2013, New York University Libraries completed a strategic plan to improve technology systems for archival content and special collections. New systems were implemented over the next four years to describe, request, index, and discover these materials, but the web publishing process for finding aids was one endeavor that, at the time, proved to be intractable. The Libraries' legacy method for transforming EAD files to HTML relied on an XSLT processor and a single stylesheet for markup. Any changes to the stylesheet to accommodate new material types or markup often caused the process to break. These challenges prevented archivists from embedding digital content and meeting accessibility requirements.

The project to replace the publishing service and redesign the finding aids web pages began in 2018 with a survey of similar tools in development at peer institutions. The project team considered two community projects: the ArcLight (<https://arclight.sites.stanford.edu/>) software led by Stanford Libraries and the ArchivesSpace Public User Interface (<https://archivesspace.github.io/tech-docs/architecture/public/>) initiated by Lyris with development from Hudson Molonglo, Yale University, Harvard University, and The Cherry Hill Company. Both software projects addressed the discovery and representation of archival content on the web.

In a default ArchivesSpace installation, the Public User Interface provides a public web view for selected underlying data stored in the collection management system. The nine partner archival repositories that use the New York University Libraries publishing system generate finding aids within the ArchivesSpace application using the EAD 2002 schema (<https://www.loc.gov/ead/eadschema.html>). Across the partners, five separate ArchivesSpace instances, both hosted and locally maintained, are used for collection management. Because the ArchivesSpace Public User Interface is directly associated with a single ArchivesSpace instance, it was not practical for use as a discovery tool across multiple partner repositories. Additionally, displaying data directly from the collection management system to the web does not align with New York University Libraries' established publishing workflow, which requires review before publication. The archivists expect to export a version of finding aids data from ArchivesSpace for preview and approval before publication to the web.

ArcLight is a Blacklight-based environment (<https://projectblacklight.org/>) built to support discovery and digital information delivery in archives. In 2018, the ArcLight software was at minimum viable product, or MVP, status. New York University archivists and technical staff expressed a similar concern about directly accessing the collection management system for real-time indexing and display on the web. There was a stronger preference for generating static pages in keeping with locally established library technology publishing practices.

The project team considered using direct application programming interface (API) queries to the ArchivesSpace instances to retrieve EAD or JSON files for publishing. However, not all instances had an API available, and there were security implications in extending this practice to all partner instances. Therefore, in 2020 the project team decided that the best way to improve the finding aids publishing process was to continue exporting EAD XML files from the partner archival

repositories, replace the mechanism that transformed the EAD files into HTML, make design improvements to the resulting web pages, and continue using a separate Blacklight Solr (<https://github.com/projectblacklight/blacklight>) tool for discovery. The library technology developers already had experience developing web content with Hugo (<https://gohugo.io/>), a static site generator written in Go that was selected for performance and ease of deployment. The project team decided to parse EAD content into JSON structured data to generate Hugo content files used to create HTML. The complex part would be displaying the EAD content correctly, which required a data model.

## DATA MODEL

A critical step in the redesign effort was determining which data elements from the partners' exported EAD files needed to be presented in the published finding aids so that the developers could then write software to extract, store, and transform that data into HTML to address the usability and digital archival object presentation issues mentioned above.

The archival repositories participating in the New York University finding aids publishing service create finding aids that conform to the EAD 2002 XML schema. The EAD 2002 schema is incredibly flexible, allowing archivists to describe complex and widely disparate arrangements of archival materials. That flexibility, however, presents challenges when defining software data structures and web page layouts. The project team did not want to build an overly complicated system that accommodated all of the inherent flexibility of the EAD 2002 schema. Rather, the team wanted to build a lightweight, manageable, and sustainable system that could meet the stakeholders' requirements and be easily modified to accommodate new features.

Therefore, to reduce the software and web-design complexity the project team defined a data model that captured the relevant subset of EAD 2002 schema elements and attributes used across the archival repositories. The archivists on the project team created the data model after reviewing archival description practices and the EAD 2002 data elements actively in use.

To help define the initial software data types, the developers used Zek (<https://github.com/miku/zek>), a tool that analyzes XML documents and outputs Go data structures. Starting with a small sample set of EADs, the project team worked through each data structure generated by Zek, adding tags and unique attributes to a CSV file that outlined the bundling of data types. This process of analyzing EAD XML, defining Go data types, and mapping those data types to JSON objects became the backbone of the new publishing system.

Once the preliminary data model was defined, the project team needed to determine if the data model captured all of the relevant elements, attributes, and hierarchies present in the existing finding aids. The developers created and ran EAD-element analysis tools (<https://github.com/nyudlts/ead-analysis-tools>), iterating over larger EAD sample sets which allowed the project team to identify gaps between the evolving data model and the actual data in the EAD files. The project team continued to run element analyses until the data model represented the subset of the EAD 2002 schema required to satisfy the stakeholders' needs.

The analysis tools are a set of minimal-functionality scripts created to answer the following questions about the EAD corpus:

- What is the maximum nesting level of each element?
- What attributes does the element have across all EADs in scope?

- What child elements does the element have across all EADs in scope?
- Does the EAD structure vary widely across archival repositories?
- What sequences of child elements are present across all EADs?
- What are the <dao @xlink:role> values set across all EADs?

Understanding the input to the finding aids publication pipeline was critical to the project's success. The project team developed tools to create appropriate data structures, identify test candidates for various scenarios, and identify critical parameters required for the web page design. Analyzing the element-analysis data gave the developers confidence that the EADs submitted for publication would be compatible with the new publishing infrastructure.

## DESIGN

The project team also took an iterative approach to the patron-facing design work. The discovery process began with a literature survey. In the 2012 publication “The Evolution of the Finding Aid in the United States: From Physical to Digital Document Genre,” Andrew Dillon and Ciaran Trace urge an expansion and reimagining, through user research, of the online finding aid. The authors invoke genre theory, a cross-disciplinary interpretation of a document type where the language choices, form, content, and context “reflect and shape” the practices and assumptions of the field in which it is used.<sup>1</sup> As a genre, earlier finding aids, printed on paper and presented to patrons in the presence of an archivist, manifested some of the twentieth-century understandings of archival science. For example, the archive must honor the original ordering of items within a collection, as opposed to being interpreted and grouped semantically; language should be specialized, using terms such as *extent* and *access point*; the focus of description should be at the collection level. Contemporary finding aids, now digital and online, and using the flexible format of the EAD schema, can leverage the affordances of their environment. Finding aids can offer multiple sorting and grouping options, present rich metadata at multiple levels of description, and use language and conventions that are understandable to researchers.

In 2017, Rachel Walton, in “Looking for Answers: A Usability Study of Online Finding Aid Navigation,” completed a literature review of finding aid usability studies to date, conducted her research at Princeton University, and produced recommendations for design to support users as they navigate.<sup>2</sup> Key points identified in the article were the need to clearly indicate digital objects (images, audio, or video available for immediate access online) and deliberately use inclusive language.

Tracy M. Jackson's 2011 paper, “I Want to See It: A Usability Study of Digital Content Integrated into Finding Aids,” emphasized that users scanned finding aids pages to learn immediately which collections had digitized content and which did not.<sup>3</sup> Obvious links to digital content at the item level were noted and used by users. Collection-level, consistently located language about the presence or absence of digitized content was also heeded.

The project team gathered examples of “finding aids in the wild”—peer institutions' works—and discussed features within these to emulate or avoid. Following their research phase, the project team was ready to articulate design goals.

- *Inclusive design*: The purpose of the finding aids should be understandable to all users without relying on jargon. Site design should use familiar patterns and icons to allow intuitive navigation. Web pages should be mobile-friendly and fast to load. They must be WCAG 2.0 AA (<https://www.w3.org/WAI/WCAG2AA-Conformance>) accessible.

- *Improved presentation of digital objects:* Images, audio, and video available online should be surfaced, easy to find, and easy to navigate. Each object should have its own page and URL so it can be bookmarked and cited.
- *Sense of place:* Archival contents are often arranged in complex hierarchies, and researchers must always know where they are in the hierarchy.
- *Branding:* For New York University Libraries research materials, branding should locate the user in the New York University information ecosystem. The same is true for finding aids from the two partner institutions.

The design process started with website wireframes created to show navigation and the placement of archival description elements on web pages. A larger stakeholder group of archivists and public service colleagues reviewed these and made modification requests based on their experiences working closely with patrons. The wireframes were updated and redistributed for additional feedback until the project team and the stakeholders reached a consensus. After approval, the project team requested a sample set of EADs from diverse, representative collections and began building a bare-bones HTML prototype to visualize the EAD data as a website. Ongoing conversations with stakeholder colleagues yielded real-time feedback on information architecture, which needed to account for various data structures. The archivists and the designer used the prototypes to refine the mapping of the data model to elements on the page, including providing a suitable label for each metadata field and determining the order in which the elements should appear on the page.

The HTML prototypes were also used for usability testing. Members of the project team sat with potential users, including researchers, archivists, and representatives from partner institutions to gather their impressions of the design's usability, brand consistency, and overall success. These conversations prompted adding, among other refinements, a dedicated page to request materials. A notable design challenge was the correct representation of digital objects: mapping the seven defined "role" attributes of the Digital Archival Object element to meaningful visual treatment. The project team worked together to determine precisely what each role conveyed about the availability and status of the object and how to express that appropriately through design.

### ***Design Improvements***

Figure 1 is an image of a finding aid published with the legacy publishing system, and figures 2–3 show images of a finding aid published in the new publishing system with the new design.


Figure 1. Example of a legacy finding aid.

See all finding aids in this repository

**Table of Contents**

- Descriptive Summary
- Historical/Biographical Note
- Scope and Contents note and Arrangement
- Access Points
- Administrative Information
- Container List
- ▶ Series I. Negatives
  - Series II. Interpositives
  - Series III. Duplicate Negatives

[Print / View Finding Aid as Single Page](#)



**THE TAMIMENT LIBRARY &  
ROBERT F. WAGNER LABOR ARCHIVES**

**Guide to the United Automobile Workers of America, District 65 Negatives PHOTOS.023**

Elmer Holmes Bobst Library  
70 Washington Square South  
New York, NY, 10012  
(212) 998-2630  
tamiment.wagner@library.nyu.edu

Tamiment Library and Robert F. Wagner Labor Archive  
Collection processed by Erika Gottfried, Mary Allison Farley, Katie Kirwan, Miriam Frank, Penny Kallou, Tim Donohue, Tylea Richard, Krista Hahn, Sarah Demott, 1979-1985; 1989-1991; 2008-2011. Revised by Erika Gottfried to reflect addition of materials, May 2014.  
This finding aid was produced using the Archivists' Toolkit on January 23, 2015  
Description is in English. using Describing Archives: A Content Standard  
Edited by Heather Mulliner to reflect updated administrative information , Sep 2014

---

**Descriptive Summary**

**Creator:** International Union, United Automobile, Aerospace, and Agricultural Implement Workers of America. District 65.

**Title:** United Automobile Workers of America, District 65 Negatives

**Dates [inclusive]:** 1940-1962

**Dates [bulk]:** Bulk, 1940-1959


**Abstract:** Between 1937 and 1942, District Council 65, United Automobile Workers (UAW) was at the center of a group of New York City left-wing locals organizing large numbers of warehouse and retail workers; it later expanded its organizing to include white-collar workers in publishing and universities. The collection consists of nearly 30,000 black and white negatives spanning the late 1930s through the early 1960s, with the bulk from the 1940s and 1950s, shot by members of District 65's own Camera Club, which functioned as staff photographers for the union's newspaper. The images document the everyday organizational life of District 65, its leaders and its diverse membership (including men and women of Jewish, African-American, Irish and Hispanic descent), while at the same time providing especially strong documentation generally for left and liberal causes of the 1940s-1960s, including civil rights, rank and file participation in union activities, working class leisure and recreational activities

**Quantity:** 32.25 linear feet 32 boxes and 1 folder

**Location note:** Materials in this collection are stored offsite and advance notice is required for their use. Please contact tamiment.wagner@nyu.edu at least two business days prior to visiting the library.

**Call Phrase:** PHOTOS.023

**Figure 2.** Example of a finding aid with new design.



[NYU Special Collections Search Portal](#) / [Tamiment Library and Robert F. Wagner Labor Archives](#) / [United Automobile Workers of America, District 65 Negatives](#)


## Guide to the United Automobile Workers of America, District 65 Negatives

### PHOTOS.023

---

Collection Overview

Request Materials

 View Images

---

FOR PRINTING OR SEARCHING

View This Guide all on one Page

---


CONTENTS

Series I. Negatives +

Series II. Interpositives

Series III. Duplicate Negatives

---


 THE TAMIMENT LIBRARY & ROBERT F. WAGNER LABOR ARCHIVES

**Tamiment Library and Robert F. Wagner Labor Archives**  
 Elmer Holmes Bobst Library  
 70 Washington Square South  
 2nd Floor  
 New York, NY 10012  
<https://library.nyu.edu/locations/special-collections-center/>

## United Automobile Workers of America, District 65 Negatives

CALL NUMBER  
PHOTOS.023

DATES  
1938-1982, inclusive; 1940-1962, bulk

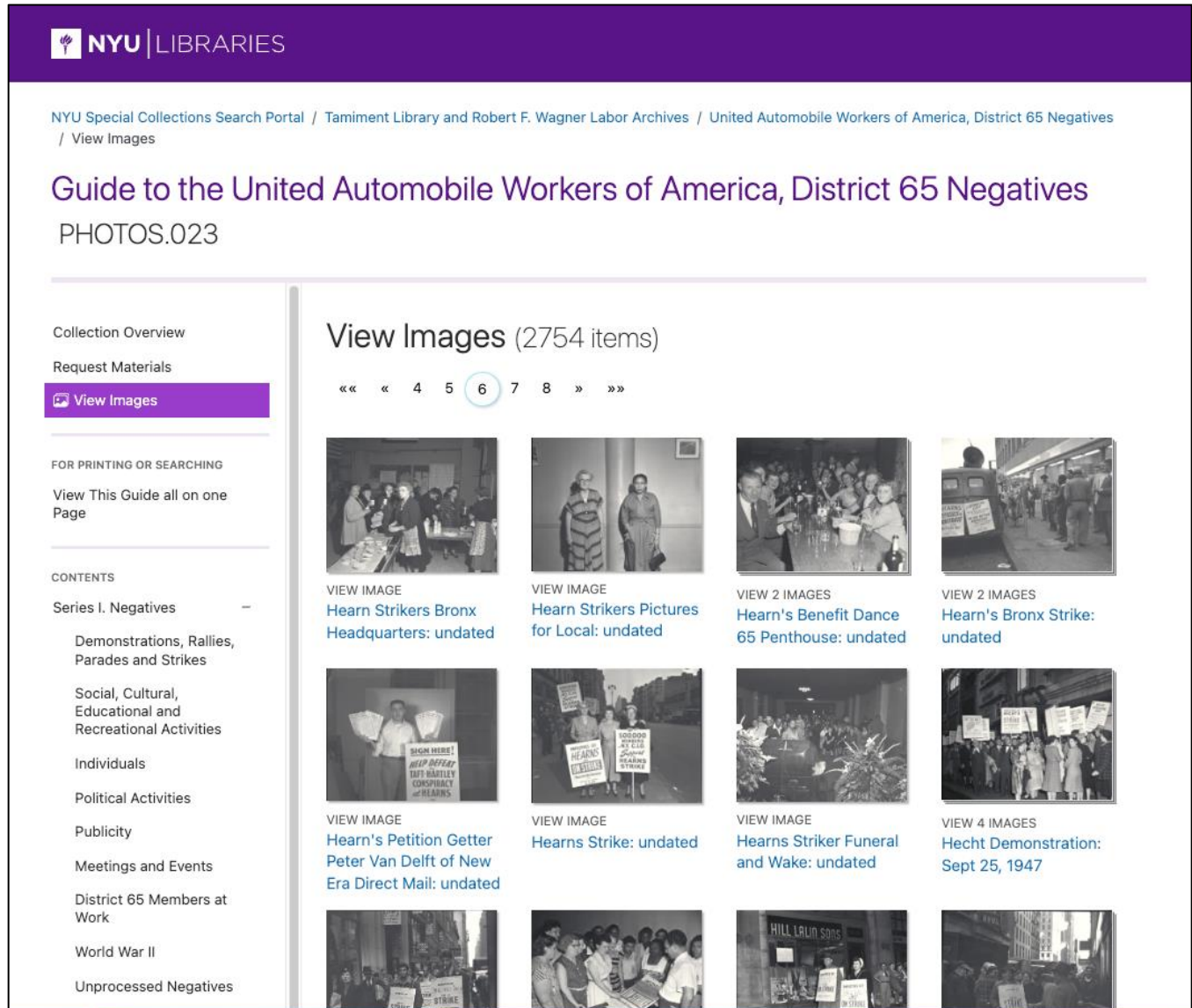
EXTENT  
32.25 Linear Feet  
32 boxes and 1 folder

LANGUAGE OF MATERIALS  
English .

ABSTRACT  
Between 1937 and 1942, District Council 65, United Automobile Workers (UAW) was at the center of a group of New York City left-wing locals organizing large numbers of warehouse and retail workers; it later expanded its organizing to include white-collar workers in publishing and universities. The collection consists of nearly 30,000 black and white negatives spanning the late 1930s through the early 1960s, with the bulk from the 1940s and 1950s, shot by members of District 65's own Camera Club, which functioned as staff photographers for the union's newspaper. The images document the everyday organizational life of District 65, its leaders and its diverse membership (including men and women of Jewish, African-American, Irish and Hispanic descent), while at the same time providing especially strong documentation generally for left and liberal causes of the 1940s-1960s, including civil rights, rank and file participation in union activities, working class leisure and recreational activities.

HISTORICAL/BIOGRAPHICAL NOTE  
District Council 65, United Automobile Workers (UAW) was organized in 1933 by a group of Jewish workers employed in dry goods warehouses on New York's Lower East Side. In 1935 it became a local of the

**Figure 3.** Example of digital objects presented in a new finding aid.



**TECHNOLOGY STACK**

The legacy finding aids publisher service was deployed in 2002. It consisted of a combination of Perl CGI scripts, Bash scripts, the Saxon Java application, and an XSLT stylesheet. An on-site Apache HTTP Server instance served the resulting finding aids HTML. In 2016, a locally maintained Blacklight instance was added to index the EADs and provide discovery services.

Although the finding aids publisher was in production for over 20 years, it had some problems. First, the HTML for all finding aids was generated using a single XSLT stylesheet. Using a single stylesheet to address the diverse needs of multiple archival repositories was difficult to achieve. Second, the developers did not possess the XSLT expertise required to maintain and update the stylesheet, making any stylesheet modifications risky. Third, the stylesheet was initially designed for EADs exported by the Archivist’s Toolkit application which was no longer in use. The Archivist’s Toolkit EADs had a different structure from the ArchivesSpace-generated EADs, requiring the development of the ArchivesSpace EAD Export plug-in

([https://github.com/NYULibraries/nyu\\_ead\\_export\\_plugin](https://github.com/NYULibraries/nyu_ead_export_plugin)) that modified the ArchivesSpace-generated EADs to be compatible with the XSLT stylesheet. The plug-in added complexity to the ArchivesSpace deployment process and meant that any changes to the ArchivesSpace generated EAD structure might require changes to the plug-in. In short, the old finding aids publication system was not sustainable.

### ***Desired Properties of the New System***

The project team wanted the new finding aids publisher implementation to have several characteristics: static site generation for publication speed and ease of maintenance, cloud deployment to align with IT-wide initiatives to move applications off site, and a desire for simple application deployment. Another requirement was eliminating the need for the custom ArchivesSpace EAD Export plug-in developed for the XSLT stylesheet.

### ***New Technology Stack***

The project team had experience with the Hugo static site generation framework and, during the technology evaluation period, learned that Hugo could be called as a library from Go programs. Developers had started using Go for other projects and found Go applications simple to deploy as single-file executables. For these reasons, the project team used Go and Hugo to create the new finding aids publisher. The new system would consist of components written in Go and deployed to Amazon Web Services (AWS). To reduce the project scope, they decided that the Blacklight discovery tool would remain and be modified as little as possible.

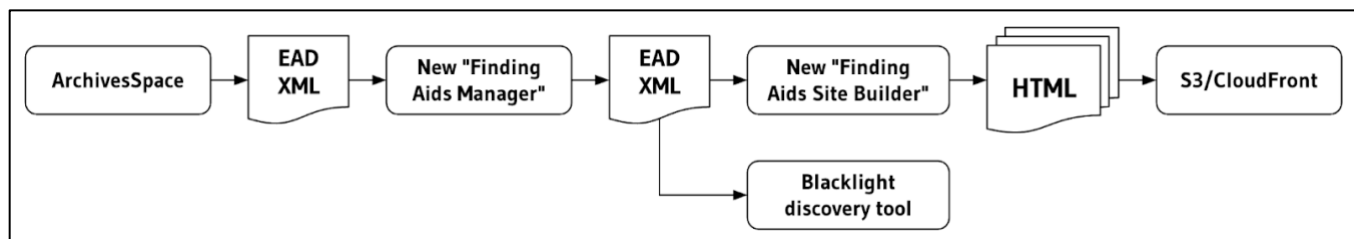
### ***Initial Concept***

The first component, the Finding Aids Manager, would replace the Perl CGI and Bash scripts used to upload, preview, publish, and delete finding aids. The Finding Aids Manager would be a web application written in Go using the Gin framework (<https://gin-gonic.com/>). The second component, the Finding Aids Site Builder, would replace the Saxon + XSLT stylesheet and would be called by the Finding Aids Manager. The Finding Aids Site Builder would generate the finding aids HTML from the incoming EADs and publish the finding aids to an AWS S3 bucket and an associated CloudFront distribution.

### ***Initial Data Pipeline***

Figure 4 is a diagram of the new system as initially envisioned.

**Figure 4.** Initial data pipeline design.



## **ITERATIVE DEVELOPMENT PROCESS AND QA**

With the general system design in mind, the project team started working iteratively on opposite ends of the data pipeline: a developer and the digital archivist worked on the data model and the EAD parsing problem while the designer worked on the Hugo templates and HTML generation. The EAD parser outputs intermediate JSON (iJSON) files used as input for the HTML-generation process. This approach allowed team members to iterate independently: the designer could work

with the latest stable set of iJSON files while the developer and digital archivist continued to refine the data model and update the Go data structures used for parsing. When a new version of the data model and parsing code was ready, a new set of iJSON files was generated and passed on to the designer.

To facilitate rapid prototyping, the designer developed scripts using the jq (<https://jqlang.github.io/jq/>) command-line JSON processor to transform the iJSON files into Hugo-compatible JSON (hJSON) files. This allowed the designer to quickly refine the Hugo templates because they could modify the jq scripts themselves instead of waiting for others.

These data modeling, parsing, and design efforts informed one another. The finding aid web page design determined which EAD elements and attributes needed to appear in the data model while the work on the data model revealed additional EAD data that might be useful in the web page design.

Although the jq scripts were useful for the prototyping phase, they were too slow to run in production. Therefore, once the iJSON and hJSON structures were mostly stable, the jq-script functionality was folded into the Finding Aids Site Builder. This resulted in a performance improvement of 96.8% for the worst-case EAD (1,142.67 seconds to 36.54 seconds) and 96.4% on average (5.01 seconds to 0.18 seconds).

### ***Nightly and Weekly Builds with a New Set of Scripts***

Once the various software components were mostly feature-complete, the developers implemented regular builds of all EADs. Nightly builds generated HTML finding aids for all EADs and pushed them to an AWS S3 bucket and password-protected AWS CloudFront distribution. Weekly builds incorporated a new set of scripts providing an EAD → HTML validation step.

Building and deploying every night provided the project team with a rapid feedback mechanism for design changes, identified build problems early, and gave the team the confidence to deploy to production when the time came.

The weekly EAD → HTML validation step made sure that all the expected data from the EADs was present in the finding aid HTML. The EAD → HTML validator identified subtle EAD parsing and data display issues. For example, the EAD → HTML validator uncovered an error in the EAD parsing code: the parser looked for elements named `<altformavailable>` instead of the correct element name `<altformavail>`. This discovery prompted a programmatic review of the EAD parsing library to ensure that there were not any additional typographical errors.

### ***Selective Stream Parsing***

As work progressed, the designer noticed that in some cases, the order of the data in the iJSON did not match the order of the elements in the source EAD XML, resulting in incorrect data in the finding aids. For example, in one `<bioghist>` element, there were child `<p>` and `<list>` elements, with the `<p>` elements preceding the `<list>`. The designer noticed that in the iJSON, the `<list>` elements appeared before the `<p>` elements. The project team realized that the choice of parsing strategies caused this. Instead of stream parsing, where an XML document is processed line-by-line, the preliminary code used DOM parsing, which loads the entire XML document into a data structure and does not preserve the order of elements in the XML document. When the project team discovered this, they were far into the development process, and it would have been challenging to change over to full-stream parsing without significantly impacting the project

schedule. They determined which portions of the finding aid required order preservation and the digital archivist developed code that implemented stream parsing only for those elements.

### ADAPTING TO CHANGE

As one might expect with a project of this magnitude, some feature development took longer than expected. The project team wrestled with three competing project constraints: scope, schedule, and product quality. The team did not want to compromise on the quality of the software or the resulting finding aids, so they revisited the project scope.

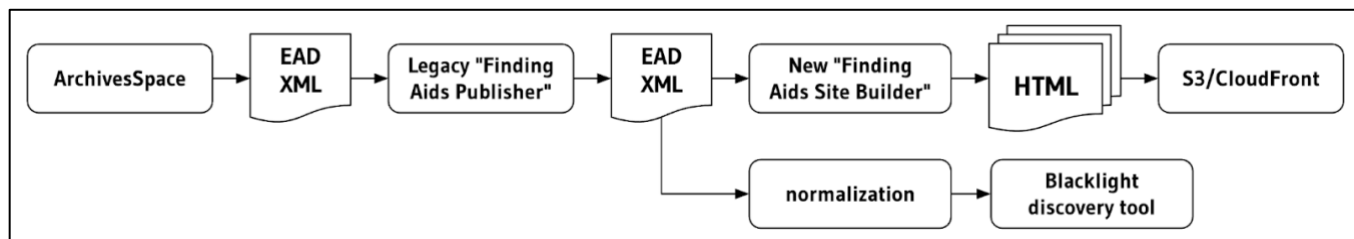
As mentioned above, the initial concept was to build a Finding Aids Manager application to replace the existing EAD publisher application, call Hugo as a library as part of the Finding Aids Site Builder, and deploy everything to AWS. To reduce the project scope, the project team decided to retrofit the existing EAD publisher application to use the new parsing and finding aid generation code and to call Hugo as a standalone application. Lastly, the project team reevaluated the deployment strategy.

Retrofitting the existing EAD Publisher made migrating the application to AWS unnecessary. AWS would be used only to serve the published finding aids via the S3 bucket and associated CloudFront distribution. The project team saved weeks and possibly months of additional work by re-scoping the project.

### Data Pipeline as Deployed

Figure 5 is a diagram of the new system as deployed.

**Figure 5.** Data pipeline as deployed.



### LOOKING FORWARD

The new Finding Aids Publisher service was launched in the fall of 2023, and all of the New York University and partner institution finding aids were republished using the new design. The look and feel of the new finding aids were received well by public services staff and researchers working in the various archival repositories. Shortly after the service went live, developers reaped the benefits of the new infrastructure. A new upgrade to ArchivesSpace version 3.4.1 introduced a new element, which appeared in exported EAD files and the published finding aids. The project team accommodated this change by adjusting the publishing templates and data element parser to ignore the additional element as per the archivists' request. The required changes would not have been possible to perform quickly or at all with the previous publishing technology stack.

Overall, the performance and maintenance of the new publisher service have been excellent. As a result of this project, developers created a mechanism to parse EAD structured data from finding aids into the more flexible JSON data interchange format. Figures 6–9 below show data moving through the finding aids publication system.

Figure 6. A component element extracted from an EAD XML file.

```

...
<c id="aspace_ref17" level="item">
  <did>
    <unittitle>Washington Square Park, Aerial View</unittitle>
    <unitid>wsp_av_1_1_01</unitid>
    <unitid type="aspace_uri">/repositories/6/archival_objects/635128</unitid>
    <unitdate datechar="creation" normal="1907/1907" type="inclusive">1907</unitdate>
    <physdesc id="aspace_720ce09e99123aae8840dc08e9cb2092" label="General Physical Description note">black-and-white print
    (photograph), 8 x 10 inches</physdesc>
    <langmaterial><language langcode="eng">English</language>.</langmaterial>
    <container altrender="Flat Box - 12 x 15.25" ...</container>
    <dao xlink:actuate="onRequest"
    xlink:href="https://hdl.handle.net/2333.1/15dv41w7" xlink:role="image-service"
    xlink:show="new" xlink:title="Washington Square Park, Aerial View"
    xlink:type="simple">
      <daodesc>
        <p>Washington Square Park, Aerial View: 1907</p>
      </daodesc>
    </dao>
  </did>
</c>
...

```

Figure 7. The component data as it appears in the iJSON file.

```

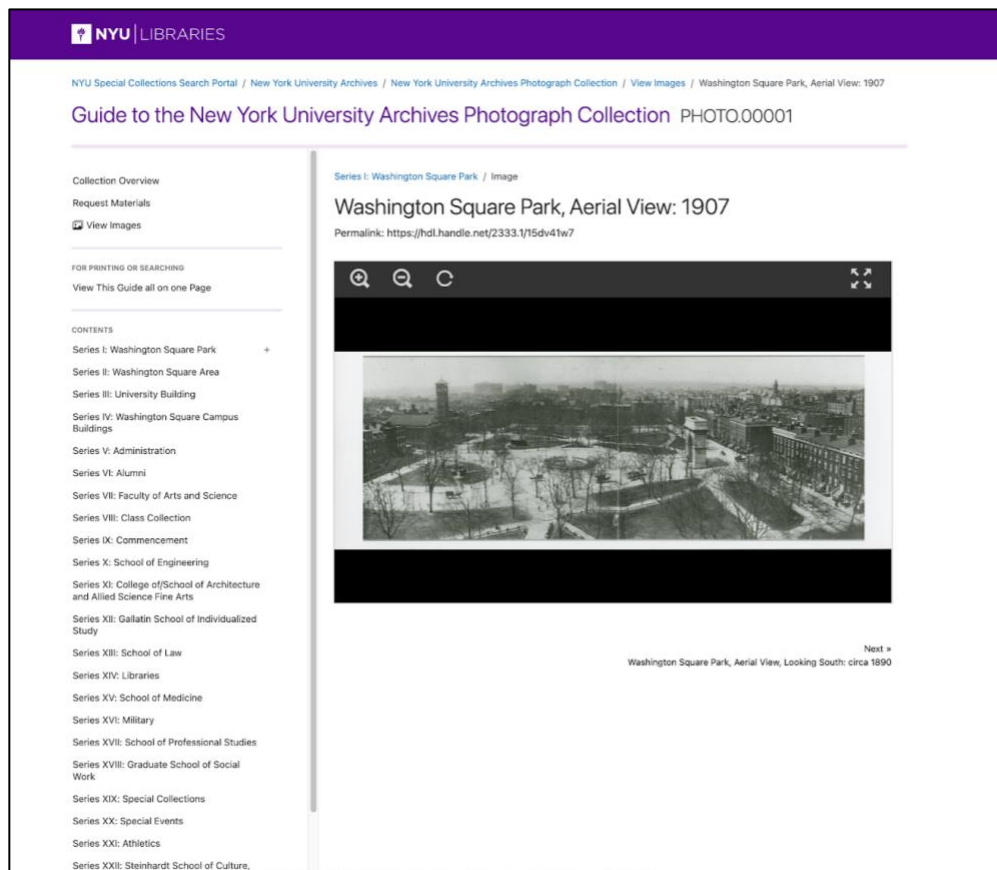
photo_00001json x
1 {
2   "runinfo": [ { 6 lines } ]
9   "pubinfo": [ { 3 lines } ]
13  "archdesc": {
14    "level": "collection",
15    "accessrestrict": [ { 15 lines } ]
31    "appraisal": [ { 39 lines } ]
71    "arrangement": [ { 255 lines } ]
327   "controlaccess": [ { 74 lines } ]
402   "custodhist": [ { 21 lines } ]
424   "did": [ { 65 lines } ]
490   "dsc": {
491     "c": [
492       {
493         "id": "aspace_ref14",
494         "level": "series",
495         "c": [
496           {
497             "id": "aspace_ref15",
498             "level": "subseries",
499             "c": [
500               {
501                 "id": "aspace_ref16",
502                 "level": "file",
503                 "c": [
504                   {
505                     "id": "aspace_ref17",
506                     "level": "item",
507                     "did": {
508                       "physdesc": [ { 6 lines } ]
509                       "container": [ { 20 lines } ]
510                       "dao": [
511                         {
512                           "actuate": "onRequest",
513                           "href": "https://hdl.handle.net/2333.1/15dv41w7",
514                           "role": "image-service",
515                           "show": "new",
516                           "do_type": "image",
517                           "count": 1,
518                           "width": 186,
519                           "height": 66,
520                           "title": "Washington Square Park, Aerial View",
521                           "type": "simple",
522                           "daodesc": {
523                             "p": [
524                               {
525                                 "value": "Washington Square Park, Aerial View: 1907"
526                               }
527                             ]
528                           }
529                         ]
530                       ]
531                     }
532                   ]
533                 }
534               ]
535             }
536           ]
537         }
538       ]
539     }
540   }
541 }

```

**Figure 8.** The hJSON file for the component's digital archival object.

```
{
  "sectionidentifier": "aspace_ref14",
  "title": "Washington Square Park, Aerial View: 1907",
  "daodesc": "Washington Square Park, Aerial View: 1907",
  "parenttitle": "Series I: Washington Square Park",
  "href": "https://hdl.handle.net/2333.1/15dv41w7",
  "noid": "15dv41w7",
  "count": 1,
  "do_type": "image",
  "role": "image-service",
  "width": 186,
  "height": 66,
  "weight": 1,
  "unitdate": [
    {
      "value": "1907",
      "type": "inclusive",
      "datechar": "creation",
      "normal": "1907/1907"
    }
  ],
  "unittitle": "Washington Square Park, Aerial View"
}
```

**Figure 9.** The component as it appears in the published finding aid.



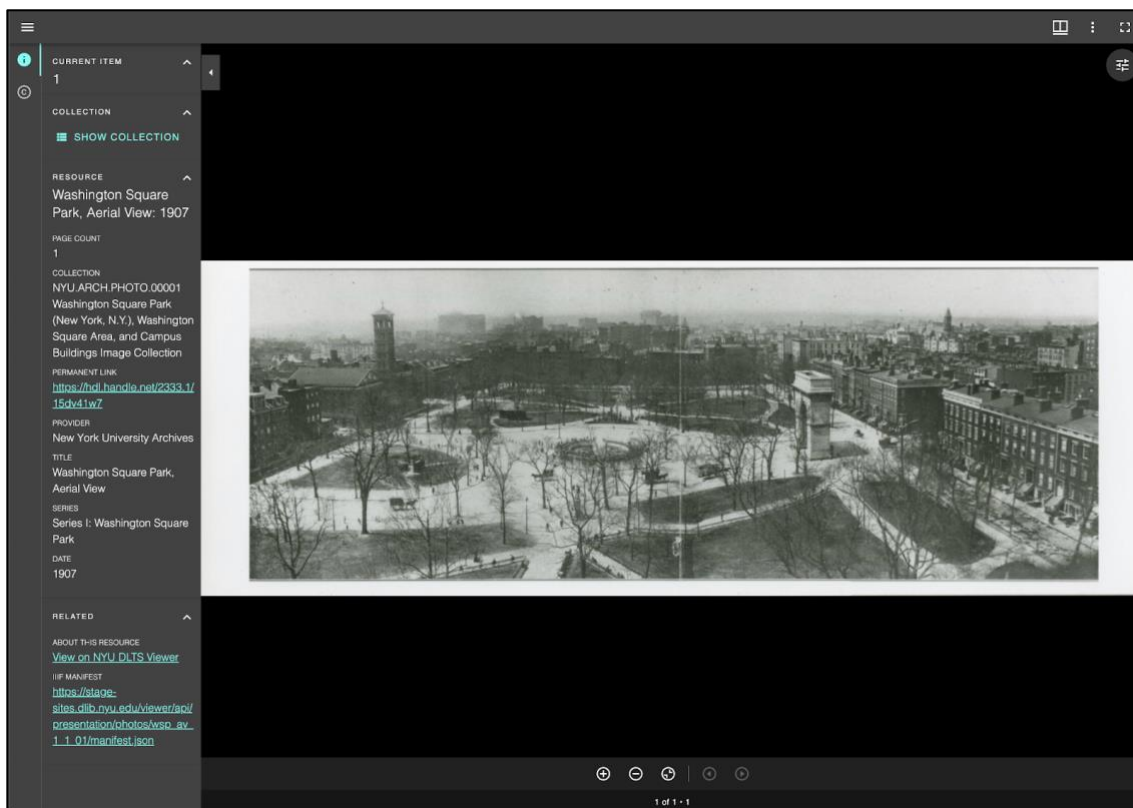
Following the finding aids project, NYU Libraries developers successfully repurposed the hJSON data to populate International Image Interoperability Framework (IIIF) manifests for digital images produced from archival collections. Figures 10–11 show how a IIIF image manifest is populated with the data from the hJSON file. Given that the iJSON and hJSON are independent representations of the EAD XML data, it should be relatively straightforward to repurpose the data for other applications.

Although the resulting software is currently in production and tailored to New York University Libraries' specific use cases, there may be opportunities for others to repurpose the code and provide suggestions or feature requests that would make the software more widely applicable. The project team is interested in hearing from other institutions to learn more about their use cases and gauge interest in using this code. If there is wider interest, it may be possible to allocate resources to refactor, restructure, and improve the code developed for this project.

**Figure 10.** An excerpt from a IIIF manifest built using the hJSON file.

```
{
  "@context": "http://iiif.io/api/presentation/3/context.json",
  "id": "https://stage-sites.dlib.nyu.edu/viewer/api/presentation/photos/wsp_av_1_1_01/manifest.json",
  "type": "Manifest",
  "label": "Washington Square Park, Aerial View: 1907",
  "behavior": [
    "paged"
  ],
  "items": [
    {
      ...
    }
  ]
}
```

**Figure 11.** The component in a IIIF image viewer using the data from the hJSON file.



---

**ADDITIONAL WORKS**

Daniels, Morgan G., and Elizabeth Yakel. "Seek and You May Find: Successful Search in Online Finding Aid Systems." *The American Archivist* 73, no. 2 (2010): 535–68.

<http://www.jstor.org/stable/23290758>.

Kenfield, Ayla Stein, and Daniel G. Tracy. "Power and Politics of User Experience: Implications of Different User Roles for Next-Gen Repository Services." *Weave: Journal of Library User Experience* 5, no. 2 (2022). <https://doi.org/10.3998/weaveux.530>.

Mayo, Dave, and Kate Bowers. "The Devil's Shoehorn: A Case Study of EAD to ArchivesSpace Migration at a Large University" *The Code4Lib Journal*, 35 (2017),

<https://journal.code4lib.org/articles/12239>.

**ENDNOTES**

<sup>1</sup> Ciaran B. Trace and Andrew Dillon, "The Evolution of the Finding Aid in the United States: From Physical to Digital Document Genre," *Archival Science* 12, no. 4 (2012): 501–19, <https://doi.org/10.1007/s10502-012-9190-5>.

<sup>2</sup> Rachel Walton, "Looking for Answers: A Usability Study of Online Finding Aid Navigation," *The American Archivist* 80 no. 1 (Spring 2017): 30, <https://doi.org/10.17723/0360-9081.80.1.30>.

<sup>3</sup> Tracy M. Jackson, "I Want to See It: A Usability Study of Digital Content Integrated into Finding Aids," (master's paper, University of North Carolina at Chapel Hill, 2011), <https://doi.org/10.17615/1hq0-ta66>.