Issues in Science and Technology Librarianship

There's an App for That

Archipelago Commons: Using the Archipelago and AMI Software to Provide Access to Rensselaer Polytechnic Institute's Engineering Drawings, A Pilot Project

Brenden J. McCarthy
Metadata Systems Librarian III
Rensselaer Libraries
Division of the Chief Information Officer
Rensselaer Polytechnic Institute
Troy, NY
mccarb2@rpi.edu

Abstract

In this paper we examine practical applications of the Archipelago Commons software for science and technology libraries through the lens of a pilot project conducted within the Libraries at Rensselaer Polytechnic Institute in 2021. This pilot project uses Archipelago for data curation, demonstrates Archipelago's value for sci-tech library applications, and highlights the technology experimentation that occurs during system migrations. The pilot involved deployment of a local installation of Archipelago and use of the Archipelago Multi-Importer (AMI) to create a digital collection of engineering drawings.

The pilot results demonstrate the value of Archipelago as a quick and lightweight "app" for any small sci-tech library that wants to showcase their digital collections or needs to quickly replace a legacy system. After this pilot we used the same approach detailed in this paper to migrate all 4,168 images from our old database to Archipelago. We conclude that Archipelago has value to sci-tech libraries and can be scaled to GLAM institutions of any size, with data as simple or robust as necessary for an institution's local needs.

Keywords: Archipelago, Digital collections, Library software, Digital asset management, Engineering drawings

Recommended citation:

McCarthy, B. J. (2022). Archipelago Commons: Using the Archipelago and AMI software to provide access to Rensselaer Polytechnic Institute's engineering drawings, a pilot project. *Issues in Science and Technology Librarianship, 101*.
https://doi.org/10.29173/istl2717

Introduction

There is a new lightweight open-source software app for managing your digital assets named Archipelago Commons, commonly called Archipelago. This paper details how Rensselaer Polytechnic Institute used Archipelago in a way that is relevant to sci-tech libraries by doing a pilot project with Library Information Services' engineering drawings, which belong to the Institute Archives and Special Collections. We used the Archipelago MultiObject Importer (AMI) to ingest this collection into Archipelago from simple spreadsheets. The goal in writing about this project is to spread awareness of Archipelago and its many potential applications for the sci-tech library community.

Background

Purpose

As a sci-tech library that serves a small R1, most of our budget is dedicated to maintaining our subscription packages from vendors to support the high research activities of our faculty and students. As a result, there are few remaining funds available to allocate towards our important local digital collections, which also have research value.

Our Archives' image database, Inmagic Genie, had reached its end of life and needed to be replaced. We ended vendor support in 2020 and the server would need to be shut down by the fall of 2022. We needed a fast, easy and free (a lofty goal) solution to display our images without a disruption in service. As part of our Digital Repository Committee's work in selecting a new system, we decided to pilot a small image collection using a local instance of Archipelago in the spring of 2021.

About Archipelago

Archipelago Commons is an open-source software application developed by the Metropolitan New York Library Council (METRO). It was conceived three years ago to lower technological barriers as well as the work involved in maintaining them (Pino Navarro & Lund, 2021). The system was designed with empathy, local institutional needs, and metadata at the heart of its development (Ziegler, 2022). Archipelago's first mention in the literature appears in 2019 during a technology evaluation conducted by the University of Toronto Libraries (Babcock et al., 2020). A demo site is available, which goes into further detail and allows users to experience the software's features (Metropolitan New York Library Council, 2022). Rensselaer Libraries began using Archipelago during Beta 3 on Drupal 8; we are now using Release Candidate 3 on Drupal 9 (Pino Navarro, 2022).

Archipelago as "App"

When we think of apps, we do not necessarily think of them by their full name, software applications—but rather, we think of the programs we quickly download and install on our smartphones to help accomplish a task. The beauty of Archipelago is its scalability. It can be used to showcase one institution's small collection, but it could be expanded to house all collections from a massive institution or consortium. It is not a heavy monolithic Digital Asset Management System or Institutional Repository, but rather a lightweight mix of integrated Drupal modules in a curated Drupal instance, running under discrete service containers (Archipelago Commons Intro, 2022). The Dockerized installation and ability to make all your digital objects using simple spreadsheets truly makes Archipelago feel like an "app" more than a "system" (though, of course, it is also the latter). In this pilot project, we used Archipelago as an "app"—a quick solution.

Using Engineering Drawings for the Pilot

As the first technological research university in the United States, Rensselaer has many digital assets from throughout its history to choose from. For this pilot project I chose to focus on our collection of engineering drawings. I selected this collection both to see if our massive TIFF files of these large-scale drawings would "break" the system (they didn't), and if Archipelago could meet our presentation and metadata needs (it did).

Method

Installation

Our pilot was conducted using Archipelago Release Candidate 1 on Drupal 8 (Pino Navarro, 2020). We were later able to replicate the pilot using both RC2 and RC3 on D9.

Our Systems Engineer followed the Docker deployment README on Github for installation (Pino Navarro, 2022). By the end of the night, we had a development server and immediately did a successful ingest through the default webform interface in Drupal admin. By the end of that week, the server was added to the cluster, and we had migrated the entire collection with enhanced metadata and Linked Open Data (LoD). Our Systems Engineer has been able to keep future releases up to date all the way from Beta 3 to RC 3 (current release) with few issues.

We needed to do a fresh installation once, and it was easy to repopulate the collection from the saved spreadsheets.

Files

After installation, the files need to go onto the server. The easiest way at the time of the pilot was using the command line to do this in batch, though more ways will be available in the future. Our Systems Engineer put the TIFFs on the server and provided me with the URL template, which I prepended to the filename and added to the CSV. The AMI would later call on these bitstreams.

(Meta)data Curation

The metadata step was the most time consuming. It should be noted that curating your data as was done in our project is not necessary. You can ingest your data "as is" and enhance and refine it later.

Archipelago and Metadata

Archipelago is schema-less, which means any existing metadata schema or standard that already exists can be stored in JSON. Whatever fields that are desired can be cast to display to end users with a TWIG template. I created custom TWIG templates for the Archipelago Multiobject Importer (AMI) and for our default metadata display. This was largely for testing and learning purposes and for many quick "app" users, the out-of-the-box TWIG templates provided will meet most needs.

Base Metadata

Our old system used a pseudo-Dublin Core schema. Prior to this project our Assistant Institute Archivist had exported and mapped these fields to Dublin Core and cleaned up the metadata using OpenRefine. This gave me a csv with DC metadata to work with.

Preparing the Data

For its ease of use I used Google Sheets to import the Dublin Core CSV and strip white space and perform additional metadata cleanup. From there I mapped the DC fields to the values used in the AMI TWIG template. The process is very straightforward. For example, dc.title = label, dc.description = description, dc.rights = rights, etc. Lastly, I created a filename column that contained the entities where the TIFFS on the server can be fetched from.

Adding Linked Open Data to Metadata

Linked Data is optional. Due to Archipelago's practical application of LoD, I wanted to take advantage of the software's impressive features. LoD can be added from Library of Congress, Wikidata, and more.

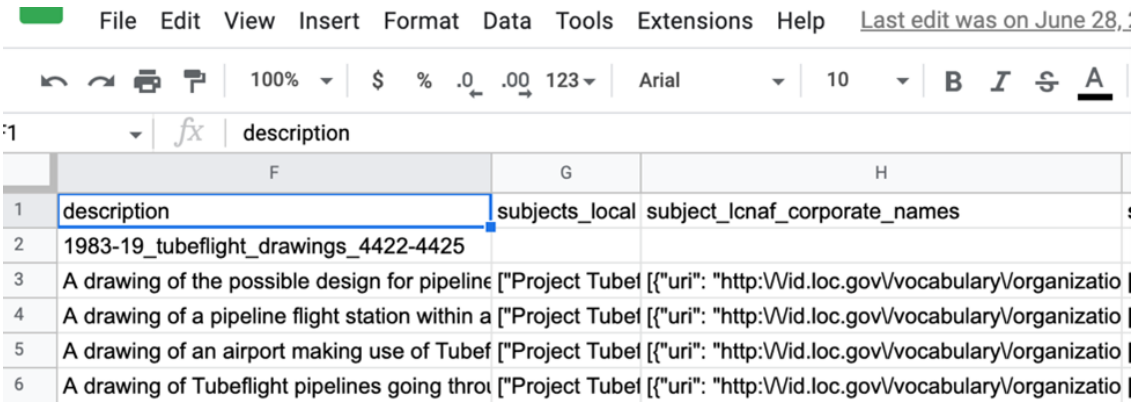

Figure 1. Spreadsheet for the AMI

Figure 1 shows a sample of some data in the spreadsheet. The description column is the Dublin Core description. The subjects_local column are local subject headings. Finally, the subject_lcnaf_corporate_names column contains the JSON for the Library of Congress Name Authority File corporate entities linked data subject headings.

I simply inserted the JSON for the LoD directly in the appropriate cell of the spreadsheet. For example, to add LoD for Library of Congress Subject Headings I created a column with the heading "subject_loc" and then added the following JSON to the cells.

```
[{"uri":"http:\/\/id.loc.gov\/authorities\/subjects\/sh85043213",
"label":"Engineering drawings"}]
```

Multiple subjects can be added and separated with a comma, for example I added:

```
[{"uri":"http:\/\/id.loc.gov\/authorities\/subjects\/sh85060743",
"label":"High speed ground
transportation"},{"uri":"http:\/\/id.loc.gov\/authorities\/subjec
ts\/sh85006607","label":"Architectural rendering"}]
```

Performing the Mass Ingest

The final step is to perform the batch import. We used AMI 0.1.0 directly from the code branch that came with the Archipelago installation. First, I exported the Google Sheet as a CSV. From there I used the Drupal administrative interface in Archipelago and navigated to Content >> AMI Sets >> Start an AMI Set. I used the spreadsheet importer option to upload the CSV and choose the AMI settings that met our needs.

Results

Within a few minutes, we had an entire public collection for our Engineering Drawings, with all our metadata present in raw JSON.

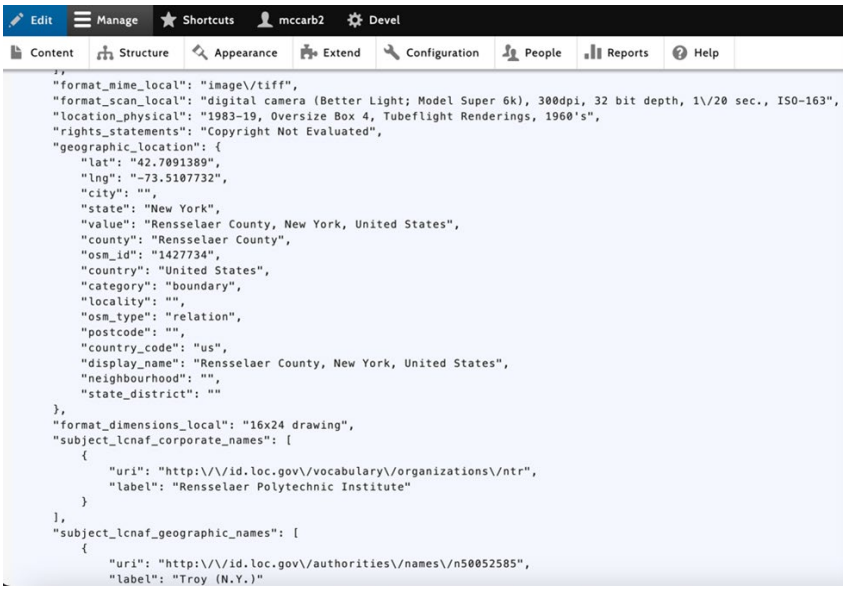

Figure 2. Raw JSON metadata in Drupal admin

Figure 2 shows our metadata from the spreadsheet, now stored in JSON and available for future use or easy export.

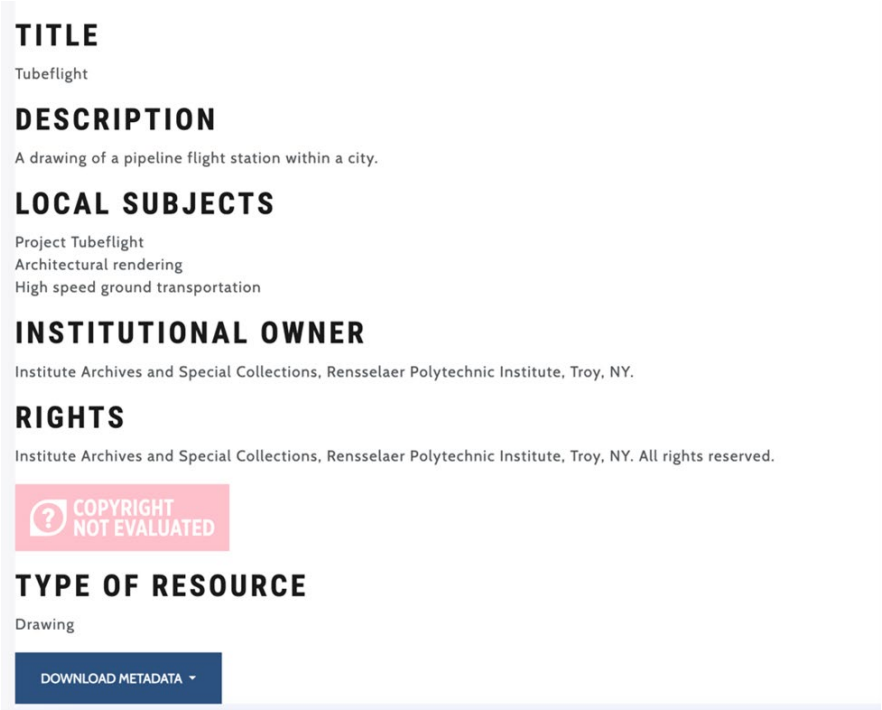

Figure 3. Public facing metadata view for a digital object

The metadata fields and labels I wanted to display to end users are controlled through the display TWIG template. Figure 3 is what a public user will see in Drupal and shows what fields I wanted to display, as well as the rights statement for the digital object.

By default, Archipelago ships with a full deck of media players and viewers. Archipelago generates an access copy of the TIFF on the fly. Both OpenSeaDragon and Mirador were tested and made excellent viewers for our engineering drawings. I am currently using OpenSeaDragon to display our engineering drawings collection, which is the system default for the type "Drawing."

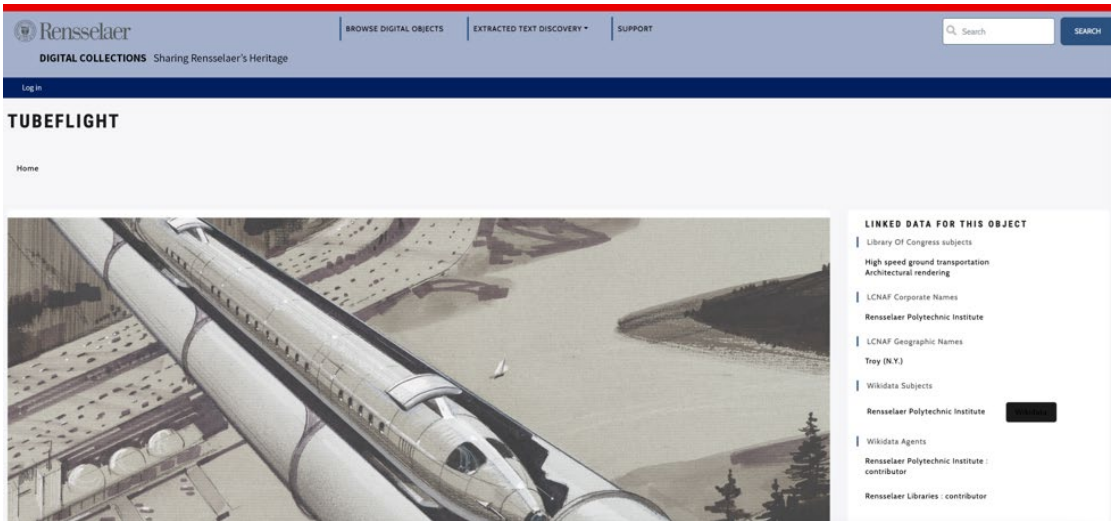

Figure 4. End-user digital object display with LoD

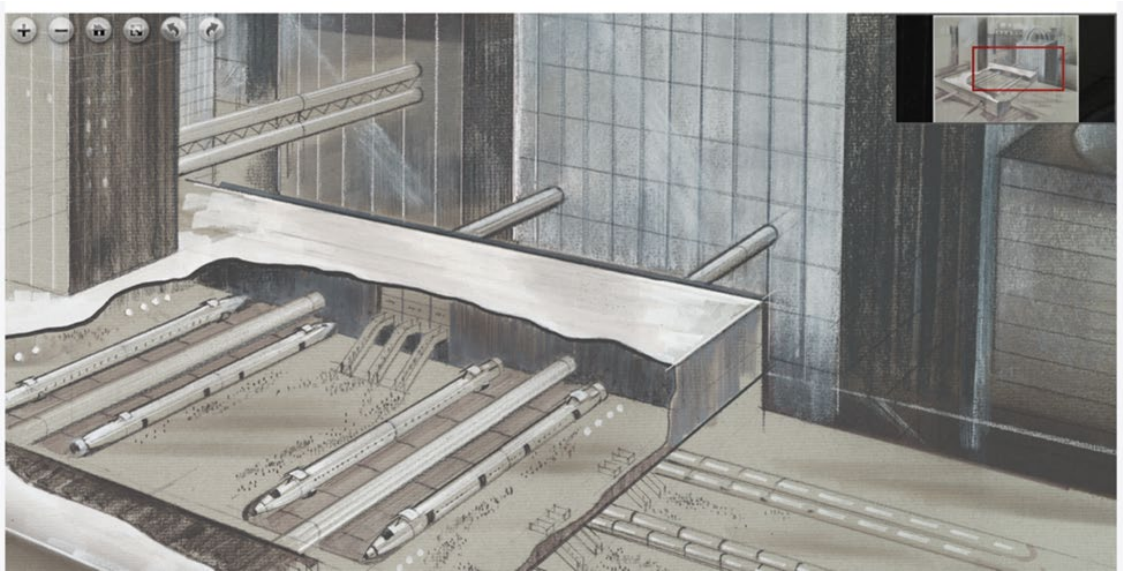

Figure 5. A built in Archipelago viewer zooming in on an engineering drawing

Figure 4 shows a digital object of a "Project Tubeflight" drawing rendered in OpenSeaDragon. Figure 4 also shows the LoD fields added on the right-hand side as well as our institutional branding at the top. Figure 5 shows another "Project Tubeflight" engineering drawing in detail using the zoom function of OpenSeaDragon.

While planned well in advance, this pilot project was completed just two days after the AMI went live. While documentation on using the AMI now exists, it was intuitive enough that I was able to do this pilot before there was public documentation available. Anyone who would like to replicate this project in the future now has robust AMI documentation available for using this mass ingest module (Archipelago Multi-Importer (AMI), 2022).

The results exceeded our expectations, and a minimally customized Archipelago deployment was far better than the existing system we had in place. Shortly after completion of the engineering drawings project, I migrated all 4,168 digital objects and their associated metadata from our old database and into Archipelago. The same approach of using the AMI and enhancing the ingestible spreadsheet with Linked Open Data was used.

Future Directions

The initiative was so successful that our Library Information Services Digital Repository Committee voted unanimously to choose Archipelago as the successor, not just for our image database, but also for our legacy digital asset management system, Digitool. This system also needed to be shut down due to security issues and we will not renew our Digitool contract with Ex Libris in August of 2022.

While the pilot project proved we are capable of hosting and maintaining our own Archipelago instance, Rensselaer Libraries have decided to contract with METRO for software as a service support and hosting in order to free up staff time. We were able to use funds freed up from our termination of legacy proprietary systems.

Conclusion

While our method may appear complicated during the first reading of this paper, I would like to underscore how accessible Archipelago is, even for the non-technical. Essentially, all that needs to be done to establish a solid digital collection is complete the dockerized deployment, place the files on the server, and then upload a spreadsheet. Doing those three things is all that is needed to go live with a collection.

There is much to be said about Archipelago that is regrettably beyond the scope of this paper, from the brilliant system architecture and design to the people— dedicated staff at METRO as well as Archipelago's ever growing user community. But to sum things up, Archipelago is empowering to the user and there is help available.

As its name suggests, you can think of Archipelago instances as a chain or cluster of islands. Some of those islands may be big. Archipelago is flexible and scalable enough that a large university could use it as their digital asset management software. Some islands might be fancy. Archipelago, being schema-less can handle any metadata schema (or multiple), whether an institution is using MODS, PBCore, DCMI, DDI, EAD, Darwin Core or something else. Some of those islands may have small islands around them—institutions can have multiple Archipelagos. For example, an institution might want one for EAD records, one for digital objects in their archives, one for institutional repository content such as Green Open Access papers and theses, and one for data set deposits.

RPI Libraries is a small island. Our pilot project demonstrates how Archipelago can be quickly used to create a new image database with Linked Open Data with only two staff members on the project. Our future direction shows how we plan to expand Archipelago into our new digital asset management system for the Rensselaer Libraries and Institute Archives and Special Collections.

I hope our experience will be useful to sci-tech libraries in similar situations and encourage them to try the demo site and then spin up their own local instance. Got digital collections, but limited staff and time? There's a free app for that and its name is Archipelago.

Acknowledgements

Special thanks to Diego Pino Navarro, Director of Digital Strategy and Allison Lund, Digital Projects and Metadata Librarian at Metropolitan New York Library Council for their enthusiasm for my work during this project. And especially to my Division of the Chief Information Officer colleague, Systems Engineer George Biggar, without whom our local deployments of Archipelago would not have been possible.

References

Archipelago Commons Intro. (2022). Archipelago. https://docs.archipelago.nyc/1.0.0-RC3/

Archipelago Multi-Importer (AMI). (2022). Archipelago. https://docs.archipelago.nyc/1.0.0-RC3/ami_index/

Babcock, K., Lee, S., Rajakumar, J., & Wagner, A. (2020). Where do we go from here: A review of technology solutions for providing access to digital collections. *Code4Lib Journal, 47*. https://journal.code4lib.org/articles/15000

Metropolitan New York Library Council. (2022). *Welcome to Archipelago.nyc.* https://archipelago.nyc/

Pino Navarro, D. (2020). *Archipelago docker deployment* (Version 1.0.0-RC1). Github. https://github.com/esmero/archipelago-deployment/tree/1.0.0-RC1

Pino Navarro, D. (2022). *Archipelago docker deployment for Drupal 9* (Version 1.0.0-RC3). GitHub. https://github.com/esmero/archipelago-deployment/tree/1.0.0-RC3

Pino Navarro, D., & Lund, A. (2021, June 7-10). *Two years of navigating the calm waters of Archipelago: A tale of multi ended openness, community nourishing and sandy beaches* [Conference presentation]. 16th International Open Repositories Conference. https://archipelago.nyc/index.php/do/22d24a63-62e1-44fc-bb20-898466e8a649

Ziegler, S. (2022). Toward empathetic digital repositories: An interview with Diego Pino Navarro. *Journal of Critical Digital Librarianship, 2*(1), 1. https://digitalcommons.lsu.edu/jcdl/vol2/iss1/1

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