

We Built a Repository!

A Migration Case Study at a Small(er) Liberal Arts College

Nora Egloff, Digital Repository Librarian

Adam Malantonio, Digital Initiatives Developer

Samvera Connect, Washington University St. Louis

24 October, 2019



LAFAYETTE
COLLEGE



a heart-warming story!



WE BUILT A REPOSITORY

a triumph!

LAFAYETTE
COLLEGE

Introduction

- A case study of our migration/merger of our Institutional Repository and Digital Collections into a single Hyrax application
 - named Spot internally, but really known as the Lafayette Digital Repository
- Since time is limited, we're just going to be covering the process of moving and remediating our objects
 - please feel free to get in touch with any questions; we could talk for hours!
- Still ongoing
 - Preview / beta went live in mid-October
 - Migration has been split into two phases; we're just beginning the second
 - Aiming to be fully migrated by Fall 2020
- new-ldr.lafayette.edu
- github.com/LafayetteCollegeLibraries/spot

About Lafayette College



* Designed by TownMapsUSA.com



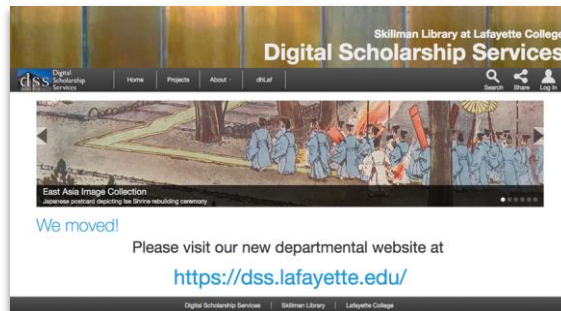
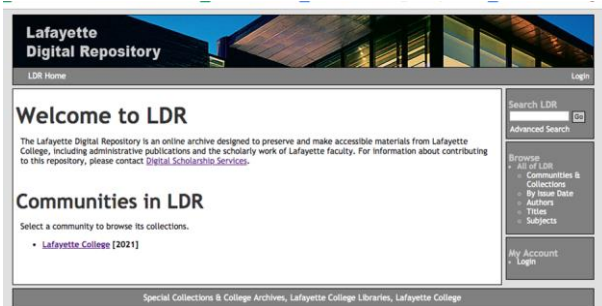
Boston Public Library, [Tichnor Brothers Postcard Collection](#)

- Located in Easton, PA in the Lehigh Valley
- Private, founded in 1826
- Total enrollment of 2,642 students
- Undergraduate
- Liberal Arts
 - Humanities
 - Social Sciences
 - Natural Sciences
 - Engineering

LAFAYETTE
COLLEGE

About the migration

What we're moving from



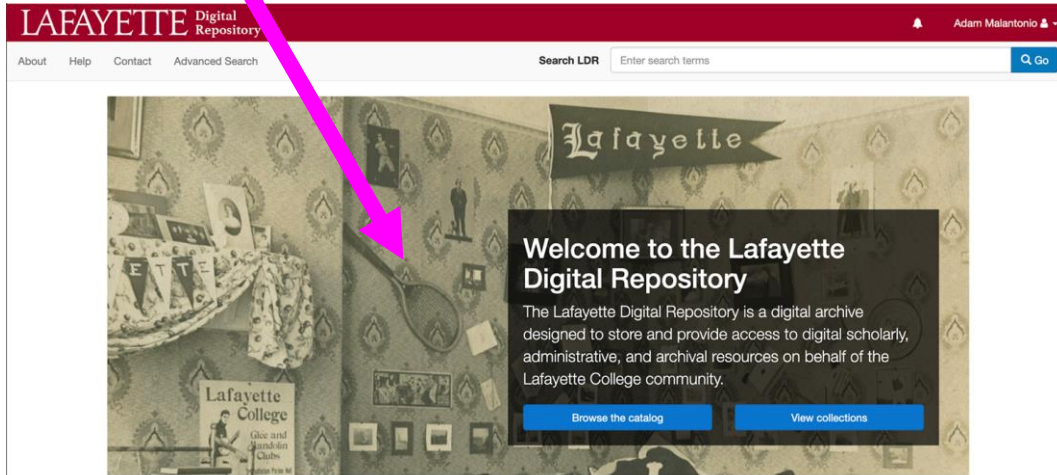
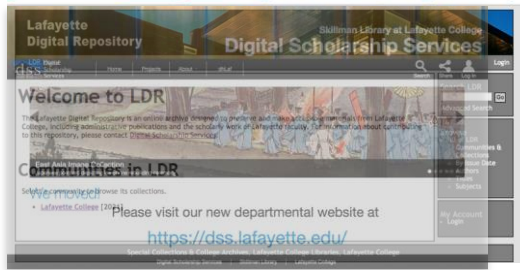
ldr.lafayette.edu (DSpace)

- ~2,000 items
- Institutional repository content
 - Faculty publications
 - Campus reports
 - Administrative Content
 - A student publication

digital.lafayette.edu (Islandora)

- ~15,000 items
- Digital collections content
 - East Asia Image Collection
 - Lafayette publications
 - Shakespeare Bulletin Archive
- Metadata edited using homegrown application

What we're moving to



- Single repository for institutional content and digital collections
- More comprehensive user experience
- One fewer individual system to manage
- (It's a Hyrax!)

What our migration plan looks like

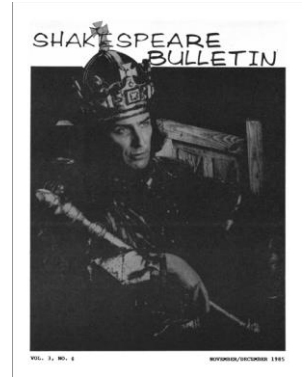
- Gather requirements and desires from library and faculty stakeholders (Winter 2017)
 - Consulted with **Data Curation Experts** to help facilitate this process
- Aim to reach feature parity with existing Institutional Repository and Digital Collections within two years
 - *Lots* of time added to account for the learning curve of not only the Hyrax codebase but also a lack of experience with Ruby
 - **Digital Repository Librarian** is created and filled (Spring 2018)
 - On-boarding, Samvera camps attended, metadata remediated, and metadata model(s) created (through Autumn 2018)
- Migrate objects with relatively simple metadata (phase 1, Autumn 2019) before moving onto the complex ones (phase 2, Spring/Summer 2020)



The image shows a screenshot of a complex spreadsheet or data table. It has multiple columns and rows, with some rows highlighted in green. The text is small and difficult to read, but it appears to be a detailed plan or inventory related to the migration project.

Phase 1: Text-Based Resources

- Institutional Repository
 - Open Access faculty publications and other scholarly content
 - College Administrative Content
 - Staff, Student, and Promotional publications
- Student Newspaper
 - September 1870-present
 - Oldest college newspaper in Pennsylvania
- Alumni Magazine
 - November 1930-present
- Shakespeare Bulletin
 - Small-press journal
- Bytes & Books
 - Library newsletter



The Lafayette

The LAFAYETTE

A L U M N U S . .



BYTES & BOOKS

Newsletter of Academic Computing Services and the College Libraries

Vol. 3, No. 1

LAFAYETTE COLLEGE

August 31, 1989

Phase 2: Image Collections

- Digitized content from Archives and Special Collections, Faculty Projects
 - Marquis de Lafayette Prints
 - East Asia Image Collection
 - College Historical Photographs
 - World War II Book of Remembrance
 - Departmental teaching collections



Pacific War Postcard Collection
Lafayette College Libraries
<http://digital.lafayette.edu/collections/eastasia/pacific-war-postcards/k0010>

Packing up for the move

Getting out of the DSpace game



- The backend is just PostgreSQL!
- Generated a CSV of metadata for all of the items
 - IR content consisted of one metadata model and was mostly maintained by two librarians
 - *chef kisses* Consistency!!
 - Eventually done with Ruby scripts + JOIN queries
- Remediated using OpenRefine and exported as another CSV
 - more on that in a minute
- Ruby scripts to generate zipped **BagIt** files (per item) for transport

The rest of the Phase 1 materials

Lafayette Newspaper, Lafayette Magazine, Shakespeare Bulletin

- These items are in our Islandora repository, **but**
 - Newspapers + Magazines digitized by vendors
 - SBA project ended up with similar directory structure
 - Metadata existed as CSVs; assets could be found in directories on library share-drives
 - asset hoarding ftw!
- So let's use the same process as DSpace content
 - github.com/LafayetteCollegeLibraries/spot-migration-tools

How about the rest of the stuff in Islandora?

- Best way to get off of Islandora was to avoid dealing with Islandora
 - Metadata edited using custom application (MetaDB)
 - enabled subject specialists to edit collections
 - allowed per collection (“project”) metadata schemas (aka nightmare fuel)
 - preservation masters stored in a directory structure that MetaDB knows about
 - Islandora essentially used as a ~~Hydra head~~ Samvera... branch? limb?
- Plan to treat it essentially the same as the Phase 1 content
 - generate CSV per collection
 - remediate in OpenRefine
 - prepare bags using same tooling

Move-in day with Darlingtonia

- Built by Data Curation Experts
 - officially deprecated :(RIP
 - versatile and flexible once you get a hang of how the pieces work together
 - better suited for batch ingesting from CSVs rather than individual packages
 - some hacking was required
- Concept of pluggable metadata mappers is perfect for per-collection metadata models
- Called via Rake task per collection/mapper type
 - `bundle exec rails spot:ingest:publication`
`source=<mapper type> path=/path/to/the/bags`
`collection_ids=<collection id>`

Assessment of metadata (Publication model)

Publication Data Model

- Work Type for Published Text-based resources
 - Shared metadata properties relating to issued date, publisher, editor, etc.
 - FileSet and file format commonalities (PDFs)
- Primarily DCMI, with some additional predicates cherry-picked from other schema/namespaces as required
- Created some local controlled vocabularies
- Questioning Authority service with the Geonames API for <dc:spatial>

Publication Data Model Metadata Application Profile

- MAP provided a working document for stakeholder review, Developer reference, and an outline for future Input Guidelines work
- For each property, MAP detailed
 - Property Label
 - Schema Mapping
 - Expected Value or Range - Controlled Source/Vocab
 - Expected Value or Range - Data Type
 - Namespace Definition
 - Obligation
 - Usage Notes
 - Source
 - Remediation Notes
 - Facetable?
 - In Hyrax Core or Basic Metadata?

Metadata Remediation - Setting Priorities

- Most metadata remediation for the Publication model was centered on improving search results and faceting user experience
 - Alternative Title property for other forms of the title (e.g. title in languages other than English).
 - Mandatory language descriptive property
 - Resource Type - make mandatory
 - Reviewed for typos and uniformity: Publisher, Source, Bibliographic Citation, Issued/Available Date
- Repository-wide metadata enhancements
 - Persistent URLs for all objects - Handle.net registry
 - Previously, only for items in DSpace repository
 - URI Rights Statements for all items in repository
 - PA Digital Rights Statement Selection Tool: bit.ly/RightsTool
 - Future work: Rights Holder evaluation

Assessment of Metadata (Image model)

Preliminary Assessment of Image Collection Metadata

- **185 metadata properties** across all Image collections
 - Removed collections deemed out-of-scope by shareholders
 - Removed all those properties showing 0.00% use, then
 - Removed technical metadata properties that will be generated during FITS file characterization
- Reviewed project-specific metadata entry guidelines and documentation; met with research assistants
- Determined that, even for metadata that was created under a given schema/namespace (e.g. Dublin Core), actual implementation varied

Image Metadata Application Profile

- Existing Digital Collections content will be mapped to an Image Work Type
 - Source materials are more varied, include:
 - art prints, ephemera and correspondence (e.g. postcards, historical teaching materials), photographs (stereocards, slides, rare picture books)
 - Metadata in MODS, sadly uninformed Dublin Core, and Things That Are Not Schemas (TTANS™)
 - Need for language-tagged title, alt title, inscription, and description properties AND display
 - Already a sizable backlog to bear in mind:
 - scrapbooks, Persian manuscripts, letters, and accruals to existing collections
 - Sought active feedback and input from stakeholders
 - College Archivist and Rare Books Cataloger
 - Faculty and research assistants

Collection-level Metadata

Collection Metadata Application Profile

- Out of the box, Hyrax Collection-level metadata is essentially the same as Generic Work Type
 - Possible area for future SMIG WG?
 - Added properties for Sponsor, Research Help Contact
 - Changed some labels to be more Collection-specific
 - Removed some administrative and rights metadata that we chose to not apply at this level

Filtering by:

LDR Type > Collection



What's next?

What's next?

- Phased metadata remediation and ingest of Image Model Collections
 - Expected through Spring 2020
- User testing (and subsequent bug fixes, UI/UX tweaking)
- Create a Metadata Entry Guidelines document for backend users
- Tackle backlog ingest of digital collections created by various departments during the migration hiatus
- Audiovisual Work Type and collections development
- DPLA dreams
 - Blacklight OAI-PMH gem
- Hey campus, LDR's not dead!
 - Faculty, Students, other Library departments, Administration, Communications Division

What we learned in the process

Community Support as an Invisible Coworker

- Community documentation, ongoing project outputs, interest group calls, and Slack channels were invaluable resources to a lone librarian tackling this work
- Shoutouts to:
 - [Samvera Metadata Interest Group](#)
 - [Hyrax 2.0 Metadata Application Profile](#) SMIG work
 - [Predicate Decision Tree](#)
 - [MODS to RDF Mapping Recommendations](#)
 - [Samvera Knowledge Base](#)
 - [DLF Metadata Support Group](#) Slack channel and in-person meeting
 - [Linked Open Vocabularies](#) project

Lessons Learned



NBC

- Importance of Hug-Ops
- Community is essential when working with a small team
- It's going to take some time! Slap on another 6 months to your estimate
 - Unforeseen delays caused us to do a soft, beta launch instead of a "big reveal"
- The Joy of Learning Project Management the hard way
 - Campus ITS and Digital Infrastructure Team challenges
 - Infrastructure Resourcing and Timeline crosstalk
- A Hyrax does not necessarily a repository make
 - One (central!) piece of a much larger application.

Thanks!

have a peek

new-ldr.lafayette.edu

look under the hood

github.com/LafayetteCollegeLibraries/spot

get in touch

egloffn@lafayette.edu

malantoa@lafayette.edu

we're on the Samvera slack

Nora Egloff

Adam Malantonio

LAFAYETTE
COLLEGE