HYRAX: An open-source digital repository solution

Scott Kushner, Systems and Emerging Technologies Librarian

Daisy DeCoster, Library Director

Saint Peter's UNIVERSITY

- Private, Catholic, Jesuit
- Jersey City campus
- Liberal arts curriculum
- 2,500 Undergraduate students -- 800 Graduate students
- 88% NJ Residents
- Ranked third (lowest) among all of the public *and* private institutions in NJ for average student loan debt

Student Ethnicity

Hispanic/Latino -- 34%

Black or African American -- 30%

White or Caucasian -- 21%

Asian -- 8%

O'Toole Library staff

Library Director (1)
Jersey City Librarians (5)
Archivist (1)
Support staff (4)

To repository or not to repository . . .

Let's do it! 👍

- Platform for student scholarship
- Platform for faculty scholarship
- Platform for archival materials (University history)
- Enhances library's role
- Enhances discovery (OCR, Google)
- Increases collaboration with other departments
- Meets our mission and connects with information literacy "knowledge practices"
- Potential for growth

Nope 👎

- Funding not available without cutting existing subscriptions
- **No dedicated staff**
- Faculty using other solutions (ResearchGate)
- ITS support is limited
- ILS might be a simpler solution for honors theses

Saint Peter's Repository Timeline

- **2010-2011 Planning and development of DSpace repository**
  - Systems librarian built DSpace repository
  - Reference librarian served as liaison with honors program, obtained permission forms and theses
  - Archivist contributed existing digital materials and metadata
- **2012 DSpace repository goes live**
  - some small archival collections included in addition to honors theses

- **2014 First doctoral dissertations included (EdD)**

- **2016 Research into new repository solutions**
  - Wishlist: Full-text searching (OCR), image gallery options, faculty researcher pages
- **2017 Project Request to ITS**
- **2018 Development of Hyrax repository**
  - student worker training, workflow planning, scanning
- **2019 Hyrax repository goes live ?!**

How I did it...
(But where to start...?)
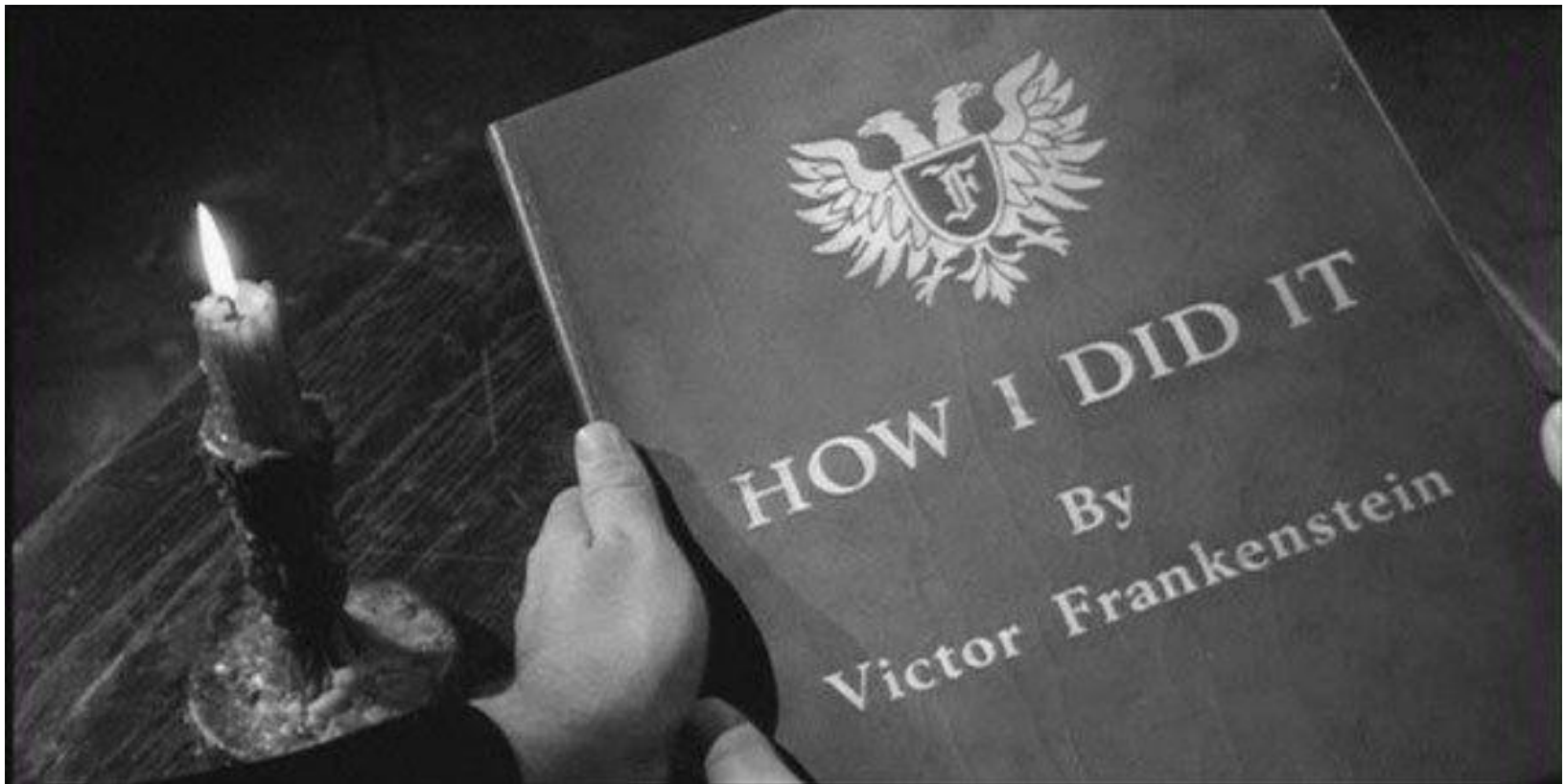
Why Hyrax.. ?

- Open Source solution for our Digital Repository software (Java, Ruby on Rails, SQL db, etc)
- Many prestigious stakeholders already invested (Stanford, Cornell, Columbia, Yale, Princeton, etc..)
- It met our technological requirements:
- ability to archive digital items
- presentation layer to display items (Blacklight, Spotlight) strategy
- Full-Text Indexing: SOLR search engine and RSOLR gem.

Where and how to implement…?

- Where will it live…? (Amazon Cloud, Heroku, our server at Saint Peter's?...somewhere else?)
- Made a pitch to IT for 2 Virtual Linux Servers on VMWare for Development
- Costs? Free? What…? (Yes there are free hosting services, each with caveats..Heroku, Amazon, etc have size, bandwidth, size, backup and restoration limitations, amongst other things.)

Hyrax requires the following software to work:

1. Solr version >= 5.x (tested up to 7.0.0)
2. Fedora Commons digital repository version >= 4.5.1 (tested up to 4.7.5)
3. A SQL RDBMS (MySQL, PostgreSQL), though note that SQLite will be used by default if you're looking to get up and running quickly
4. Redis, a key-value store
5. ImageMagick with JPEG-2000 support
6. FITS version 1.0.x (1.0.5 is known to be good, 1.1.0 is known to be bad: https://github.com/harvard-lts/fits/issues/140)
7. LibreOffice
8. ffmpeg

Actually, it's called, "Samvera"…
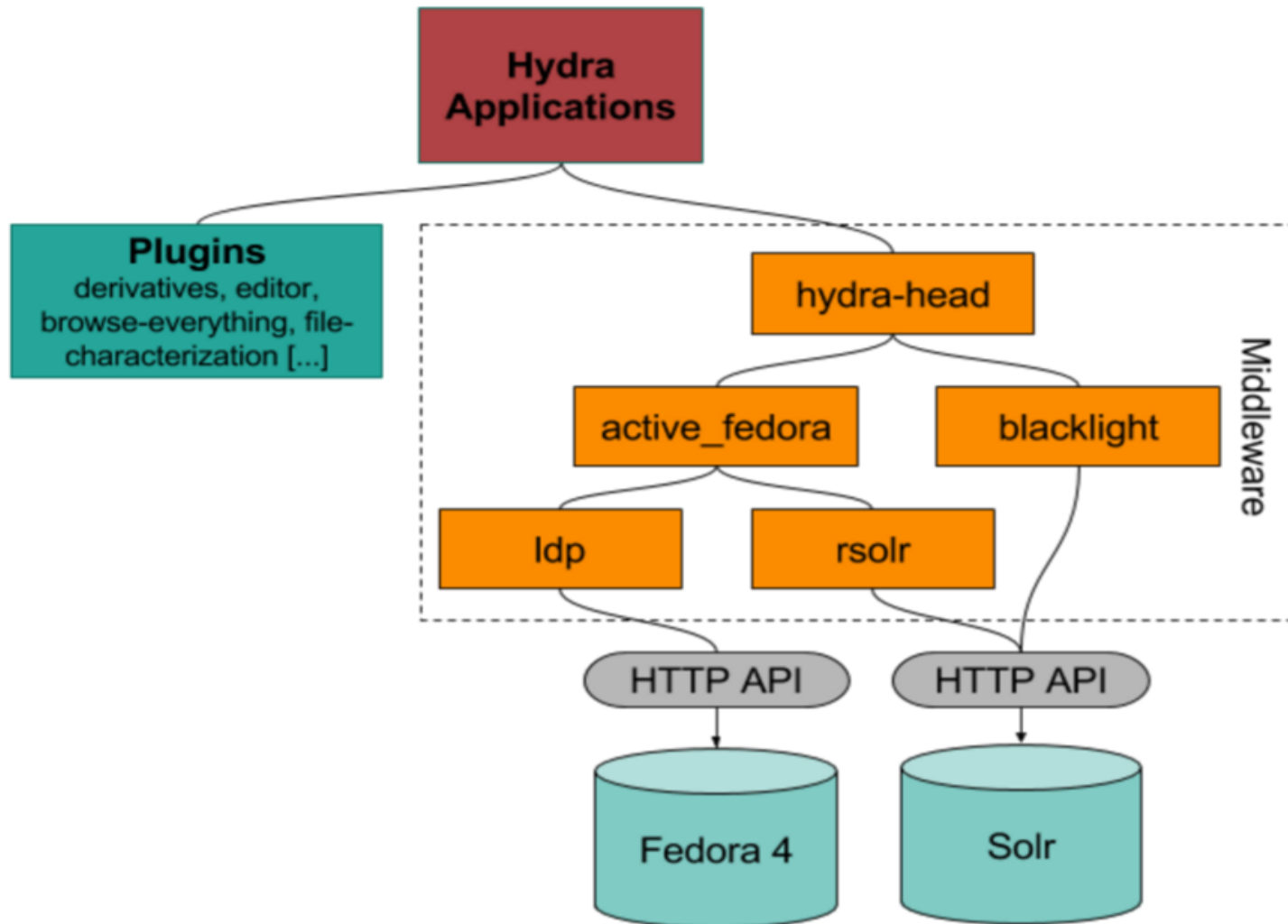
Technology Stack

Hyrax

If you are starting a new Samvera project, we strongly recommend you start with Hyrax.

At the top of our architecture diagram is "Samvera Applications." Broadly, these are Ruby on Rails based applications that follow Samvera conventions. While it is possible to create your own Samvera application by assembling the right components, and this has been a common practice in the Samvera community in the past, ongoing maintenance of home grown solutions will be more expensive than sharing maintenance costs with the rest of the community. We are in the process of focusing our community development efforts on Hyrax, and that will be the code base we are best able to support and maintain.

Fedora 4

We use Fedora 4 as our persistence layer. This is where the actual content and its associated metadata (or pointers to them) are stored. Interaction with Fedora happens via an HTTP API. Fedora 4 stores content as linked data.

Solr

We use Apache Solr for our search. Content from Fedora is indexed into solr via ActiveFedora, one of our middleware gems. Interaction with solr also happens via an HTTP API.

Middleware

hydra-head

Hydra-Head is a Ruby-on-Rails gem containing the core code for a web application using the full stack of samvera building blocks.

active_fedora

Ruby on Rails usually follows the Active Record pattern to persist objects to a database. We instead use ActiveFedora to persist objects to Fedora.

ldp

We use a ruby gem called ldp to implement the LDP (Linked Data Platform) interaction patterns for interaction with Fedora 4. You can read more about how we use LDP containers here.

rsolr

Rsolr is a ruby client for solr. Handle Rails/Solr interactions with your app.

blacklight

Much of our search and display behavior is inherited from Blacklight. Many Samvera institutions also run Blacklight applications separately from Samvera, to provide search and discovery for their collections. The Blacklight Project also has many of its own plugins, such as Spotlight, for building virtual exhibits, and GeoBlacklight, which enhances Blacklight for use with geospatial data

.Plugins

derivatives

A gem to create derivatives for uploaded content. This might include, for example, generating thumbnails for large images, down-sampled audio and video for web steaming, or thumbnail snapshots of PDF documents.

editor

Hydra-editor is a basic editor for samvera objects.

browse-everything

Browse-everything is a rails engine providing access to files in cloud storage. Currently there are drivers implemented for Dropbox, Skydrive, Google Drive, Box, and a server-side directory share.

file-characterization

hydra-file_characterization uses fits to characterize files and extract metadata about them. It might tell you what kind of image encoding an image uses, along with it's height and width for example.

Hyrax Homepage

Featured Works & Researcher

Blacklight Search with Facets...

Feature by Collection Type

Project Blacklight Spotlight

Issues...challenges..?

Due to its extensive list of dependencies, Samvera is difficult for smaller institutions to implement*

*Wikipedia

- Issues with RVM, had to use RBENV Ruby installer and Rails..

- Issues with Exec.js, therubyracer, and Rails 5.

- Nodejs installation., Yarn, etc..

- Hyrax has no frontend UI development feature (This is a Ruby on Rails thing...css and the asset pipeline, webpacker, nodejs, other stuff....)

- Better documentation (recipes) would make Hyrax easier to implement.

- Bugs and version conflicts...yes, unfortunately, even in Open Source software, there are bugs (not just in Microsoft..the good news is that they get "fixed" much faster..)

- Issues with Development vs. Production installations (web server -- Apache or NGINX, app server -- Phusion Passenger, and Capistrano...)

HELP!!!
Where to get it..
Community

Samvera

Samvera-Tech Google Group

Hyrax Github installation instruction

Hyrax Development installation instructions

Project Blacklight Spotlight Wiki

Contributing to Spotlight Wiki

Open Source Community for Samvera

Future Developments

- Valkyrie..Valkyrie is a gem which uses the Data Mapper pattern (https://en.wikipedia.org/wiki/Data_mapper_pattern) to abstract out interaction with metadata and file storage backends, restricted to the use cases of a digital repository. The goal is to enable institutions to share front-end or middleware code, but still make a choice that matches their institution's needs around where to store their metadata and files on the backend.
- Solr search result set text in context rather than just PDF attachment
- Hyrax Roadmap

The Hyrax IS cool…!

Why?

Close cousin to the Elephant!

Has tusks like the Elephant!

Looks like a Vampire!

THANK YOU, VALE MEMBERS!

Questions?

Comments?

Suggestions?

Scott Kushner -- skushner@saintpeters.edu
Daisy DeCoster -- ddecoster@saintpeters.edu, @daisydecoster