
Hercules v Hydra: Breaking Apart Curate

Jeremy Friesen

Digital Library Frameworks Specialist

University of Notre Dame

jfriesen@nd.edu

[@jeremyfriesen](#)

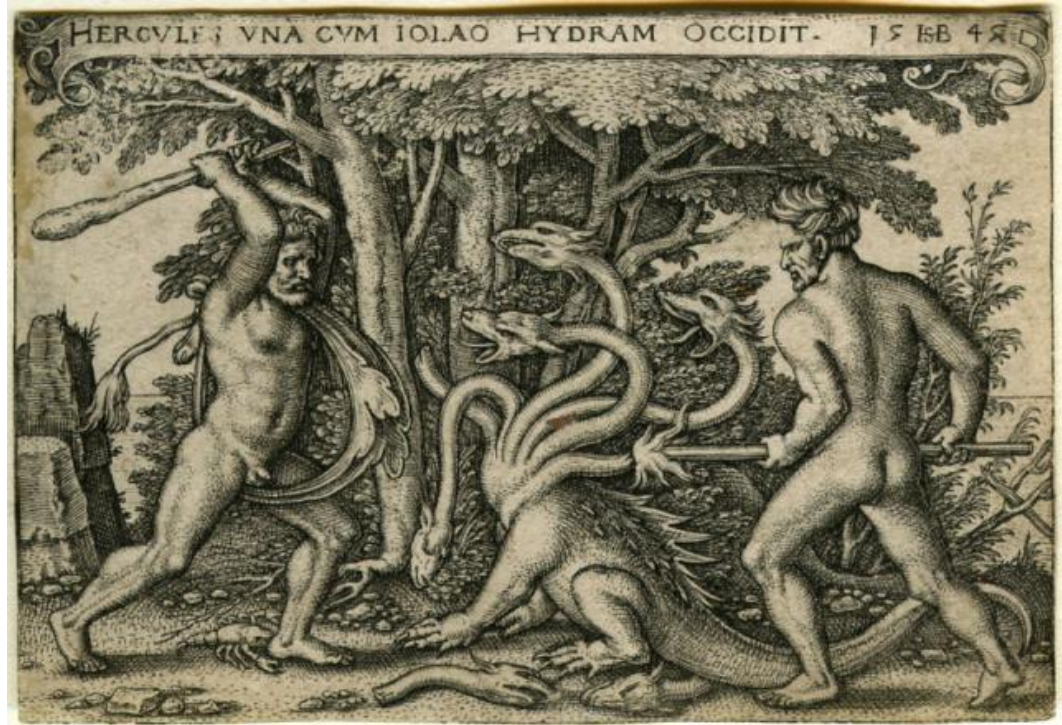
github.com/jeremyf

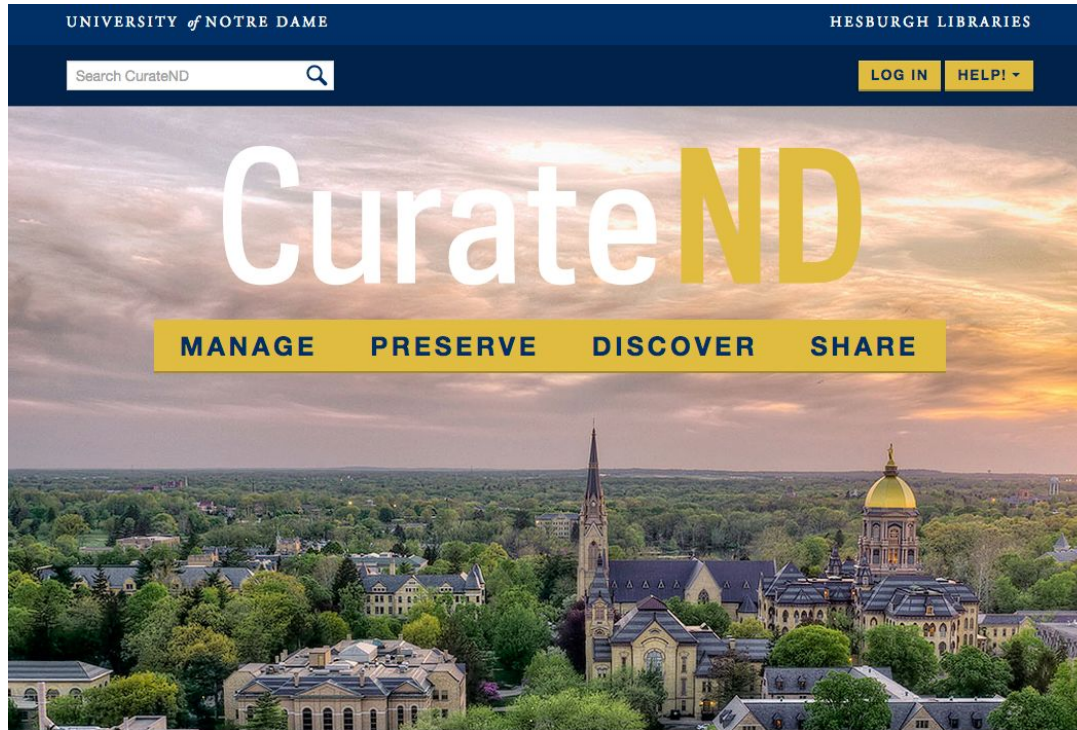
ndlib.github.io

Presentation at goo.gl/4IEQKf

First Rule of Collaboration:

Name (or don't
name?) your project
after an unkillable
mythological
creature.





Our Institutional Repository Service

- A fork of Sufia
- Models are precursors to PCDM
- User defined Collections
- “People” are stored in Fedora

CurateND

<http://git.io/vZ330>

Fedora

Data & Metadata Repository



File system

Preservation storage



Noids

PID minting service
<http://git.io/vZ33S>



MySQL

Relational data storage for applications



CAS

Campus single-sign-on



Worker Pool

Background processing; part of the CurateND codebase



Redis

Key-value store used as a work queue

Solr

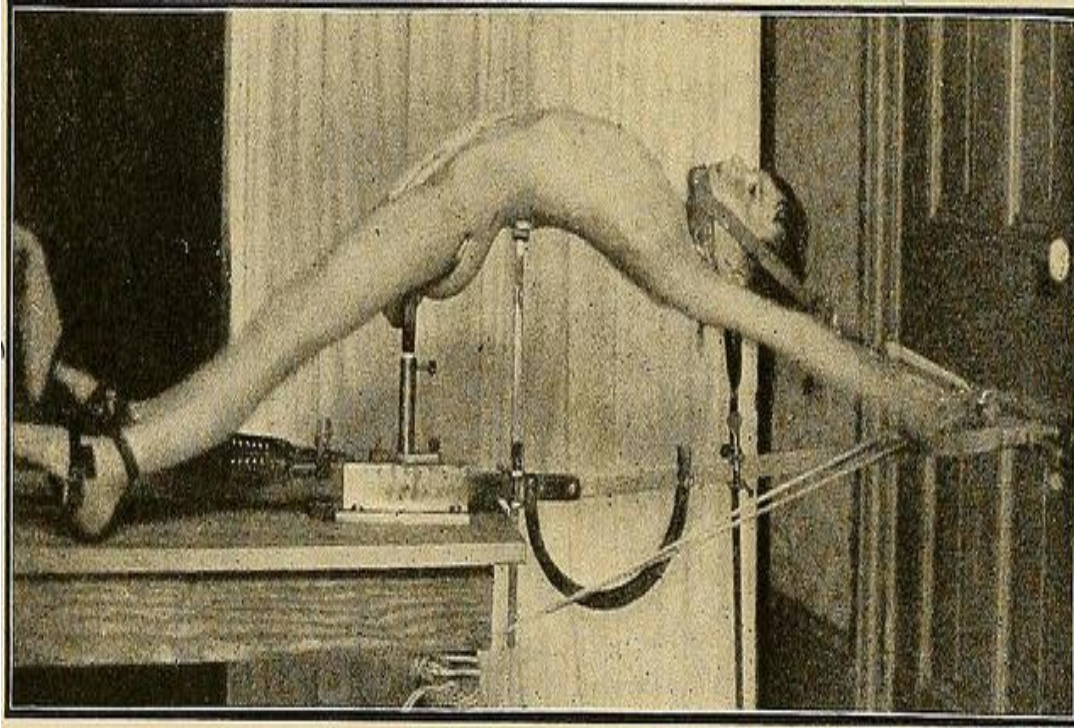
Faceted, full-text search

2014 Edition



CurateND still needs:

- Mediated Deposit
- Federated Authentication
- Group Management
- Administrative Sets™
- Batch Ingest
- More Robust Data Models



We wanted to
create something
extensible and
maintainable...we
failed to deliver
that in our
Institutional
Repository
Application

But we are
committed to
CurateND as our
Institutional
Repository *Service*



Hercules v. Hydra



So we've begun
decomposing our
monolithic application
and growing our
services



We halted development and did an inventory of what we had and what we needed by reasking the following questions:

Who, What, When, Where, Why, and How?

The People in Our Neighborhood

- Undergrads
- Graduate Students
- Researchers
- Professors
- Administrators
- Collaborators from other schools
- Lone wolf collaborators
- Metadata specialists
- Curious bystanders





Thing 1, 2, & The

- Conference seminars
- Scientific simulations
- Video captured lessons
- Retinal scans
- Master's thesis
- A book
- A kitchen sink?

Hickory Dickory Dock

- Delayed access to objects (Embargo)
- Removing future access to objects (Lease)
- People gaining and losing various responsibilities
- Format migration of objects
- Application of retention policies





Oh, the Places You'll Go

- Fedora
- Solr
- Redis
- RDBMS
- File system
- Rich web pages
- HTTP API endpoints
- Command line utilities

Why Does the Sun Shine?

- To capture the varied scholarly output of the university
- Providing a compelling reason for depositing and using the CurateND Service



How are we going to make the CurateND Institutional Repository service maintainable and extensible and answer each of those needs?

Hercules v. Hydra



Let's make Lego
Kits!

Hercules v. Hydra



But stop before we get to this point!

Hercules v. Hydra



Let the Wild
Rumpus Start

CurateND - Our Institutional Repository

At present it...

- Allows for self-deposit of rigidly defined “works”
- Manages users and account information
- Allows for arbitrary collection creation

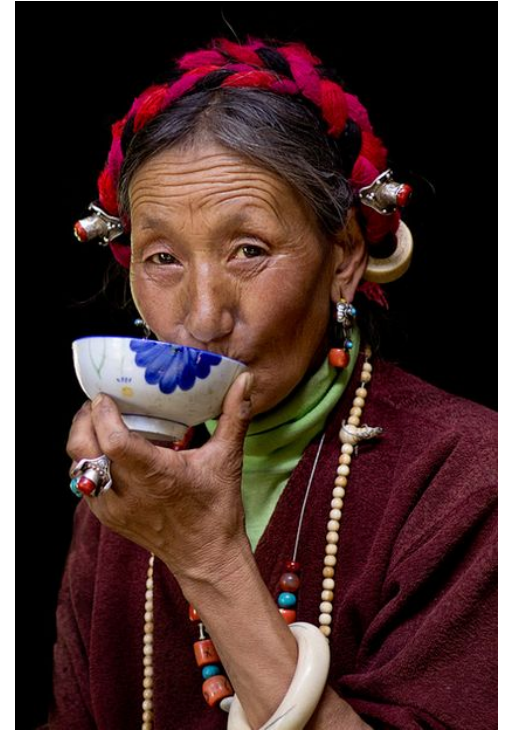
Going forward it will be a conceptual umbrella composed of other parts.

Sipity - A Patron-Oriented Workflow Application

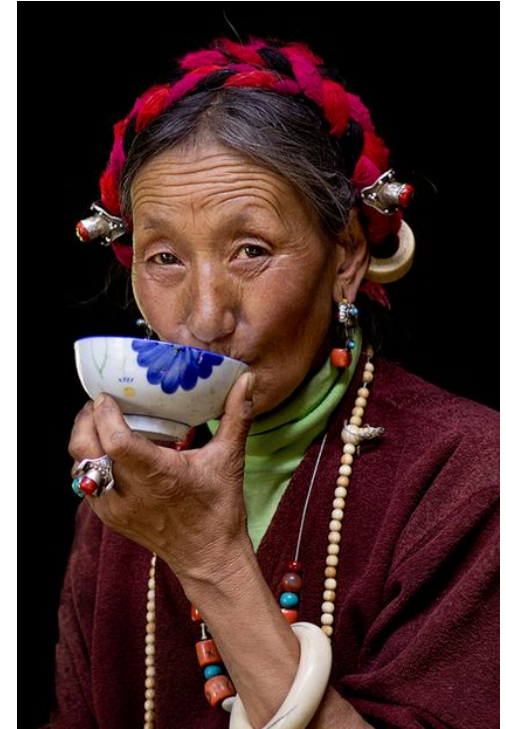
A patron-oriented workflow tool

Sipity - A Patron-Oriented Workflow

Began as a replacement for a venerable
ETD approval system written in
unmaintain(ed|able) Perl.



Sipity became a generalized approval workflow application with an initial focus on generating Submission Information Packets (SIPs).



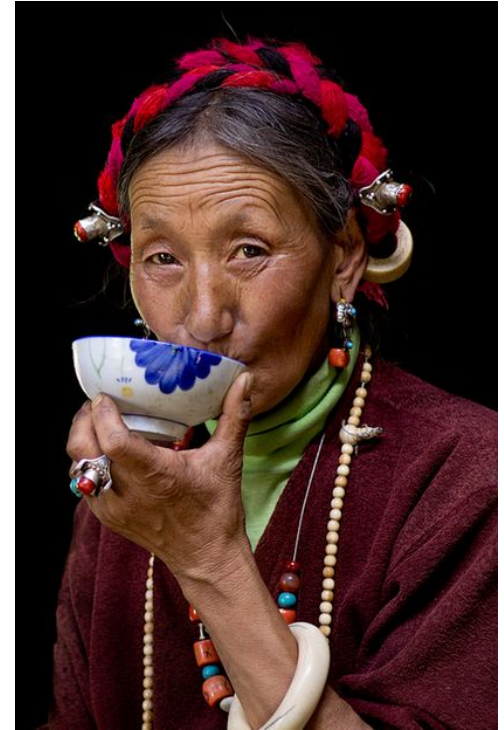
Sipity is responsible for...

- Modeling an approval workflow
- Granular user or group permissions to actions at
 - Submission Item level
 - Workflow level
- Exposing a list of Todo items at each step
- Packaging up a user submission for ingest



Sipity has no knowledge of Fedora or Solr instead focusing on the workflow.

As a result, Sipity can be leveraged for more than preparation of submission packets.



Sipity is built with a focus on extensibility through:

- Defining narrow interfaces
- Separating concepts into module spaces
 - Behavior
 - Data types
- Creating objects that model business logic





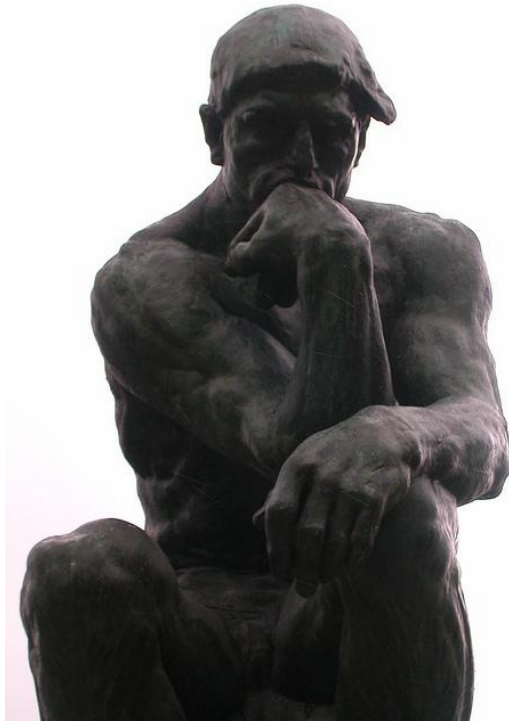
Cogitate - identity and authentication

Created as a response to standing up applications with User database tables...and always needing to account for people that were not in our LDAP service



Cogitate is responsible for...

- Translating an identifier to related identifiers and groups
- Registering groups and group members
- Exposing a single authentication end point for campus users and non-campus users
- Allowing non-Notre Dame people to be included in groups



Cogitate has no knowledge of Fedora or Solr

Cogitate focuses on aggregating identifiers (verified & unverified) for a given person (i.e. their NetID, Orcid, Twitter handle, etc).



Cogitate is NOT a Role management system, instead providing identifiers that other applications can use for permission management.

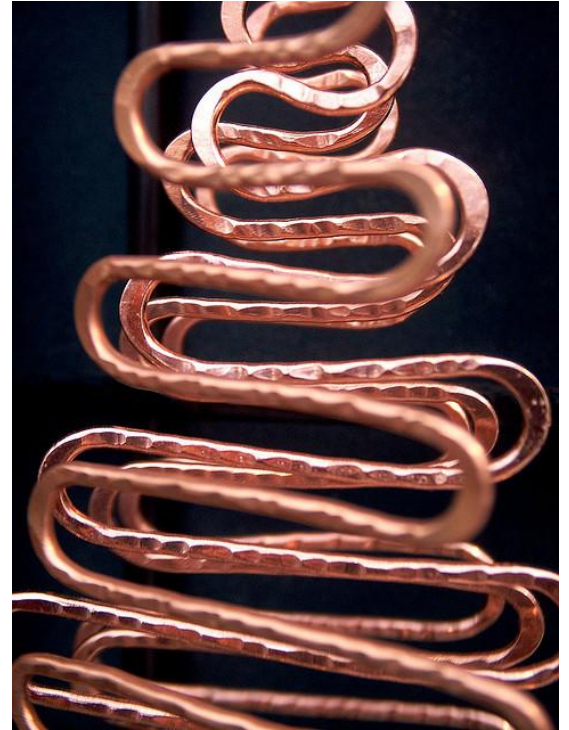


Cogitate is built with a focus on extensibility:

- Defining narrow interfaces
- Registering strategies for identifiers that are verified or unverified

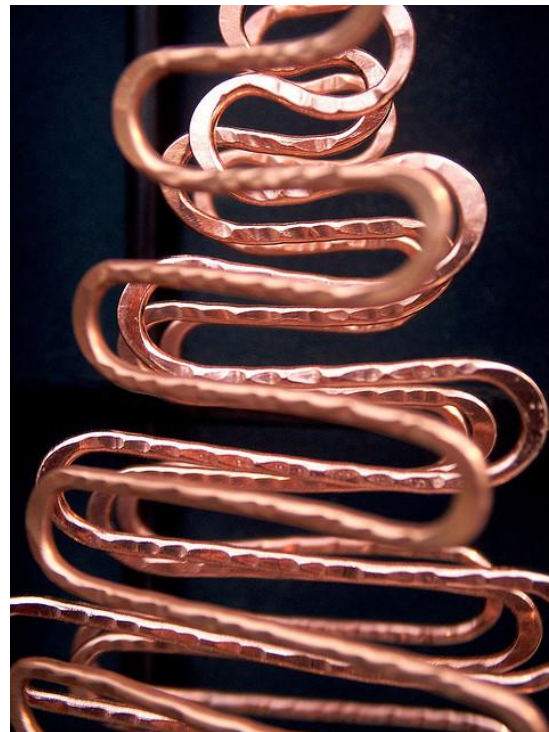
Bendo - A file preservation interface

Created as a response to Notre Dame's purchase of a tape storage system for capturing our research data.



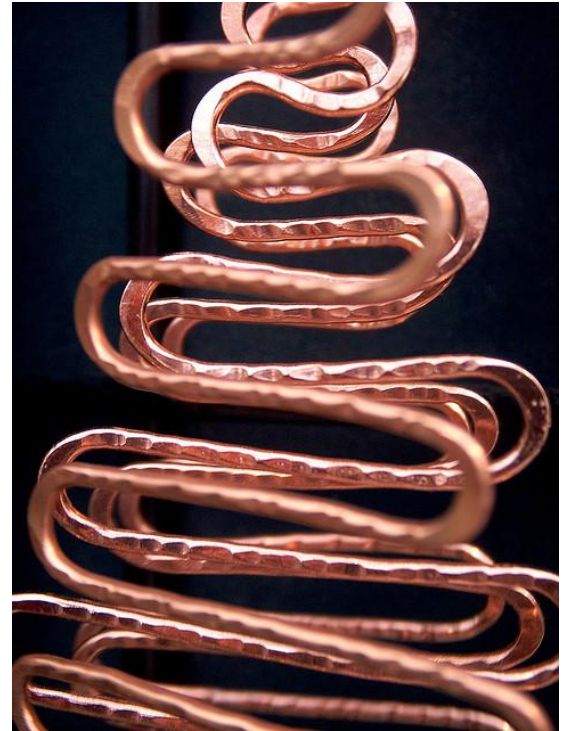
Bendo is responsible for...

- Exposing a REST endpoint for tape system
- Bundling files into larger zips to optimize tape usage
- Performing fixity checks
- Versioning content



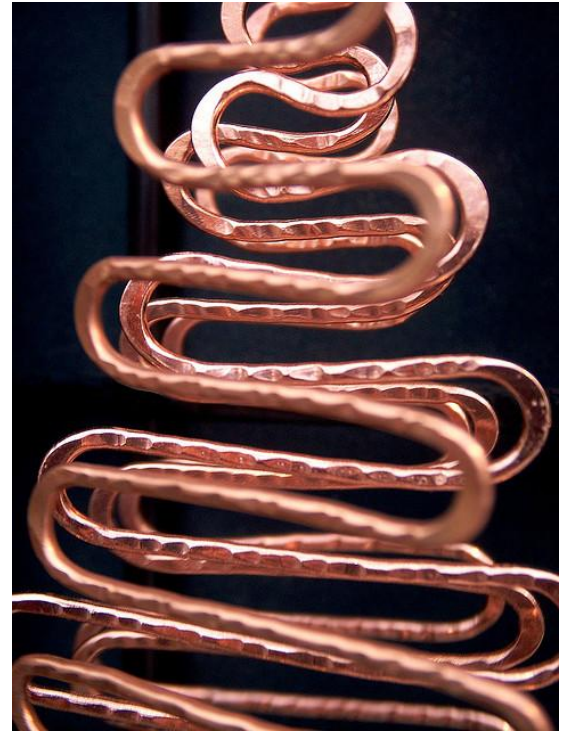
Bendo has no knowledge of Solr nor Fedora.

It focuses on negotiating the complexity of preservation by exposing a narrow interface



Bendo is built with a focus on flexibility:

It exposes a “Copy on write” behavior. You can get production data in development mode yet only write that data to the development environment.





Disadis - Fedora Download Proxy

Created out of a desire to not lock the Rails request cycle when downloading files from Fedora.



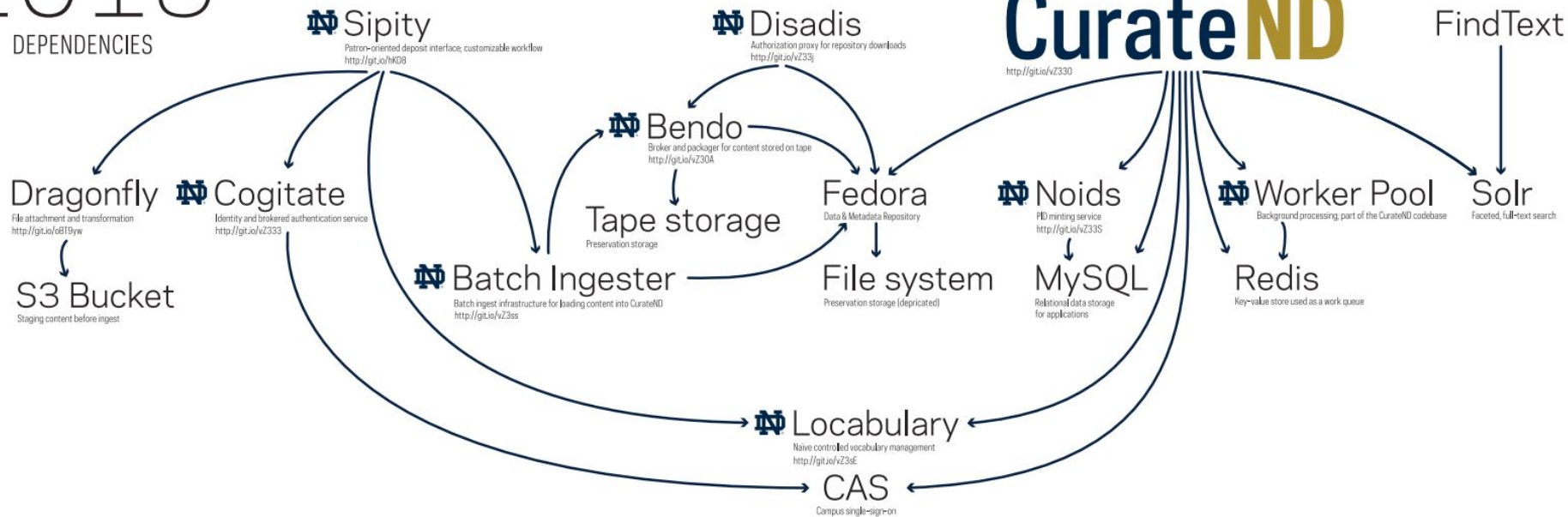
Disadis is responsible for...

- Handling the downloads on behalf of Hydra applications
- Understanding and enforcing HydraRightsMetadata
- Providing eTag and Range support
- Working with Fedora 3.6

Hercules v. Hydra

2015

DEPENDENCIES



We are focusing
on boundaries of
responsibility
across the breadth
of CurateND, our
Institutional
Repository
Service.



We are creating narrow interfaces between those boundaries...because our conjecture is that many small things are easier to test, extend, and maintain than one large thing.



Thank You

Jeremy Friesen

Digital Library Frameworks Specialist

University of Notre Dame

jfriesen@nd.edu

[@jeremyfriesen](#)

github.com/jeremyf

ndlib.github.io

Presentation at goo.gl/4IEQKf