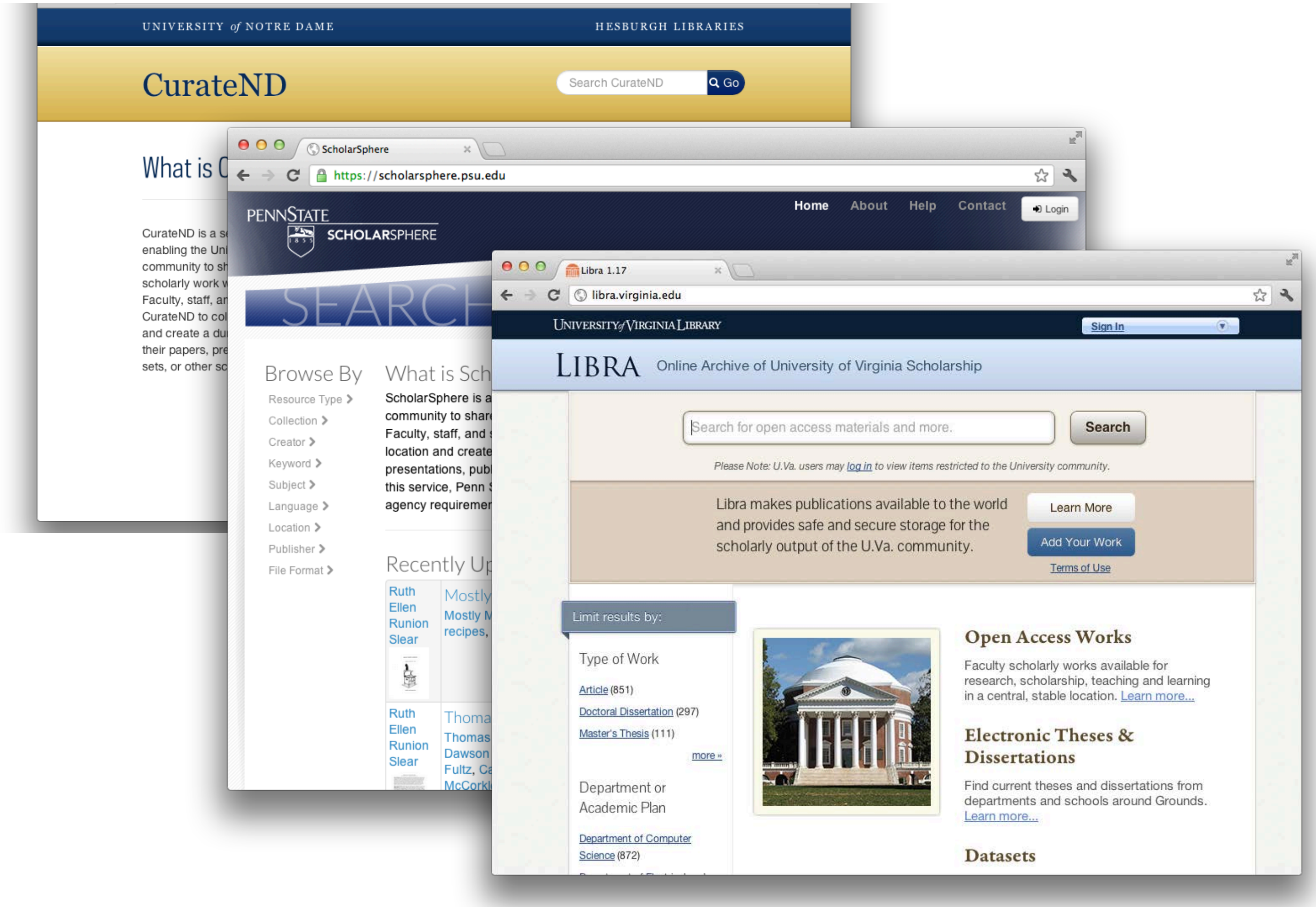
Diving into the Technology of Hydramata:

A Pluggable, Extensible Hydra Solution for Research Repositories

Rick Johnson, University of Notre Dame
Linda Newman, University of Cincinnati
Open Repositories, 2014

Why are We Here Again?



UNIVERSITY *of* NOTRE DAME HESBURGH LIBRARIES

CurateND

What is C

CurateND is a s
enabling the Uni
community to sh
scholarly work w
Faculty, staff, ar
CurateND to col
and create a du
their papers, pre
sets, or other sc

ScholarSphere

https://scholarsphere.psu.edu

PENNSTATE SCHOLARSPHERE

SEARCH

What is Sch

ScholarSphere is a
community to shar
Faculty, staff, and s
location and create
presentations, publ
this service, Penn S
agency requiremen

Recently Up

Ruth Ellen Runion Slear — Mostly — Mostly M recipes,

Ruth Ellen Runion Slear — Thoma — Thomas Dawson Fultz, Ca McCorkl

Libra 1.17

libra.virginia.edu

UNIVERSITY *of* VIRGINIA LIBRARY

LIBRA Online Archive of University of Virginia Scholarship

Please Note: U.Va. users may log in to view items restricted to the University community.

Libra makes publications available to the world and provides safe and secure storage for the scholarly output of the U.Va. community.

Limit results by:

Type of Work

Article (851)
Doctoral Dissertation (297)
Master's Thesis (111)
more »

Department or Academic Plan

Department of Computer Science (872)

Open Access Works

Faculty scholarly works available for research, scholarship, teaching and learning in a central, stable location. Learn more...

Electronic Theses & Dissertations

Find current theses and dissertations from departments and schools around Grounds. Learn more...

Datasets

New World of Scholarship

- Multimedia Student Projects
- Non-text Thesis
- Big Data
- Any Format as Data
- Everything is connected

Shared Needs

DATA CURATION EXPERTS

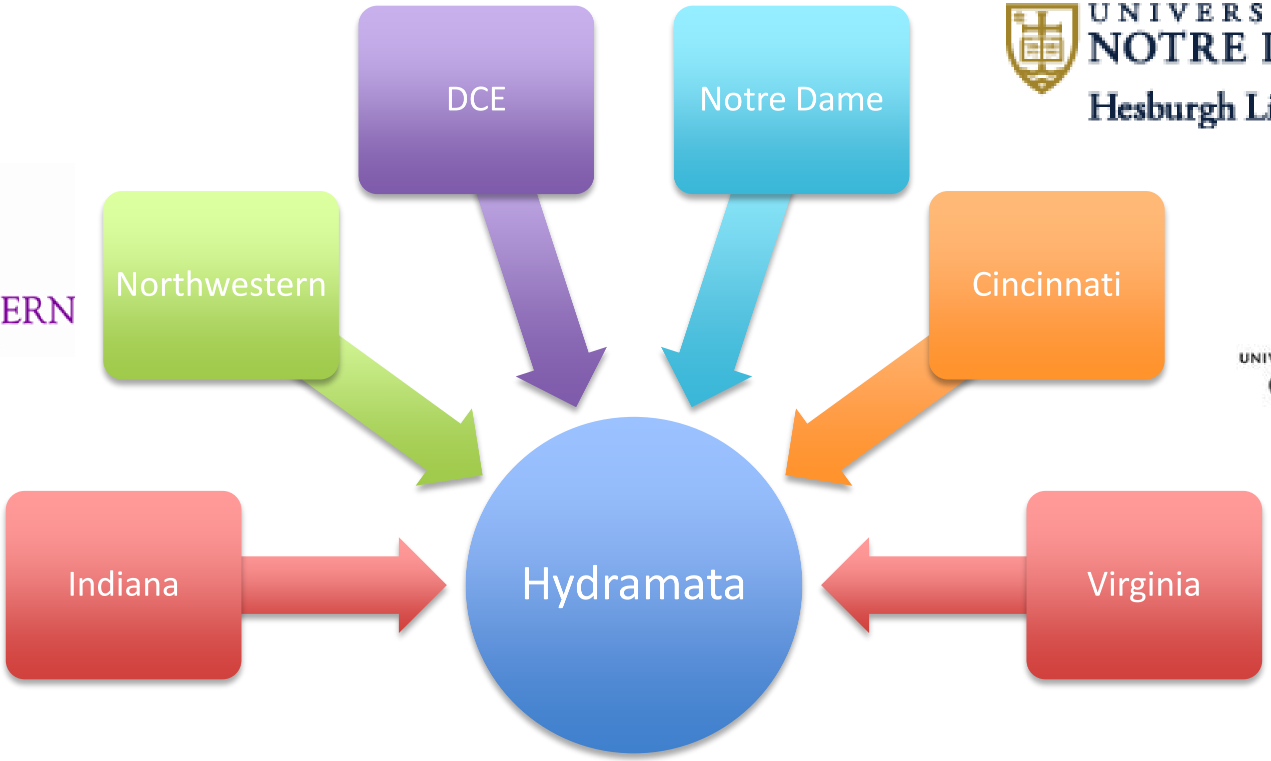
UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

NORTHWESTERN UNIVERSITY

UNIVERSITY OF Cincinnati

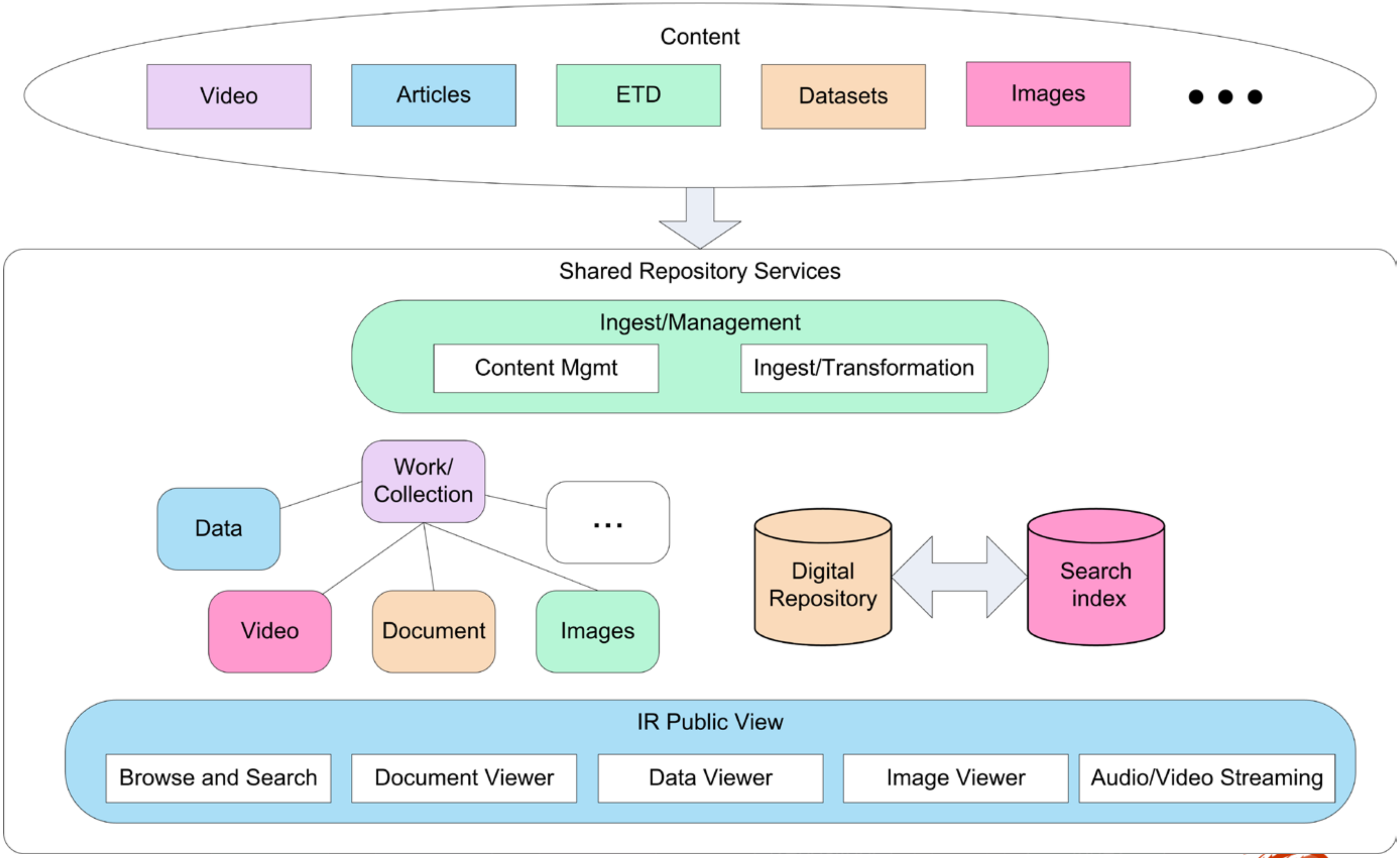
INDIANA UNIVERSITY

UNIVERSITY of VIRGINIA



…

Curate Gem Mixed Architecture



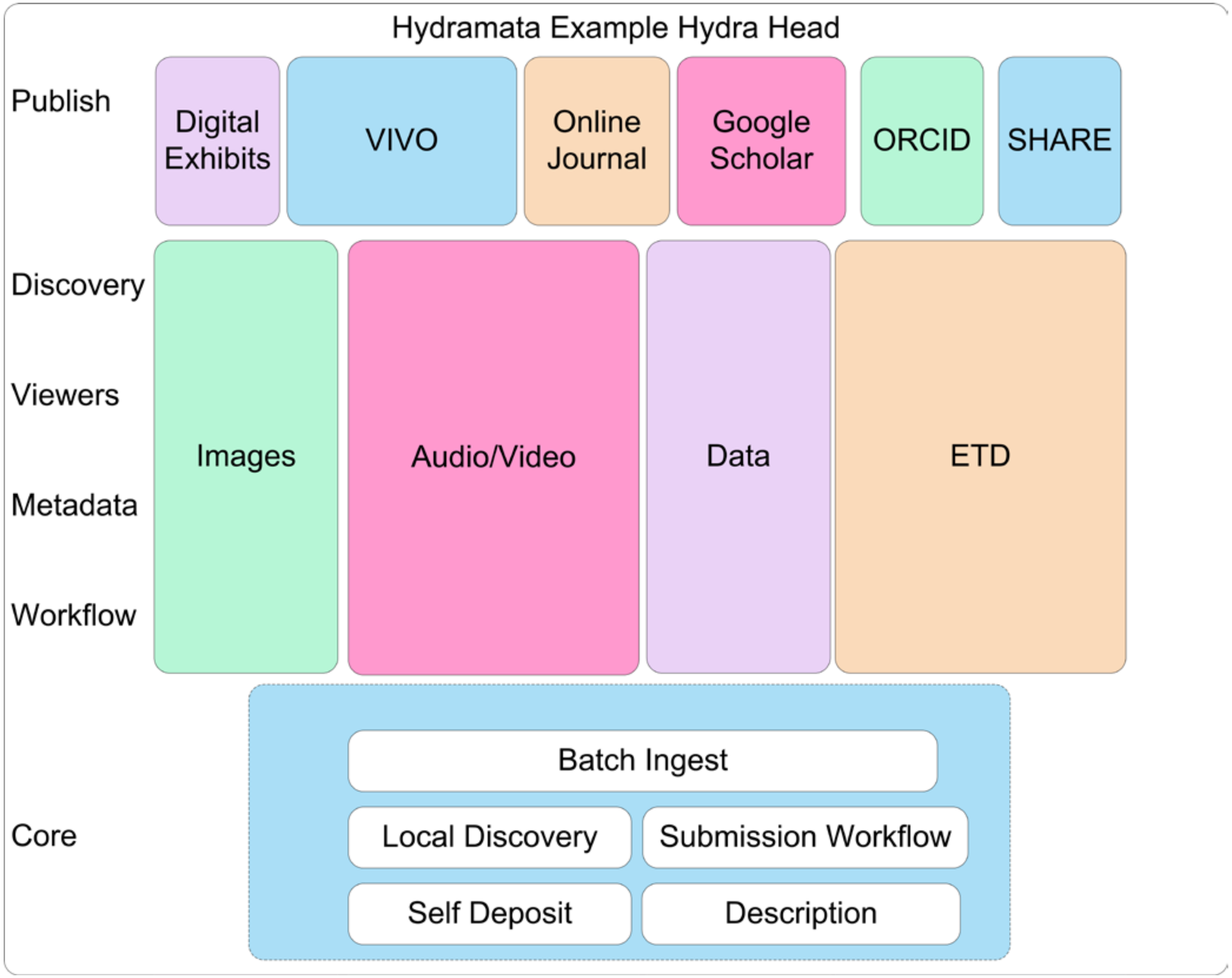
Emerging Goals / Refreshed Vision

- Each institution’s implementation is part of a larger ecosystem not merely a standalone application, and thus the code should support institutional customization.
- The core system should be separate from discrete components and provide a reference implementation for each component. Each component can then be extended or exchanged according to the needs of an individual institution.
- Component customization may also be shared among like-minded institutions because the interface between the components and the core will remain consistent.

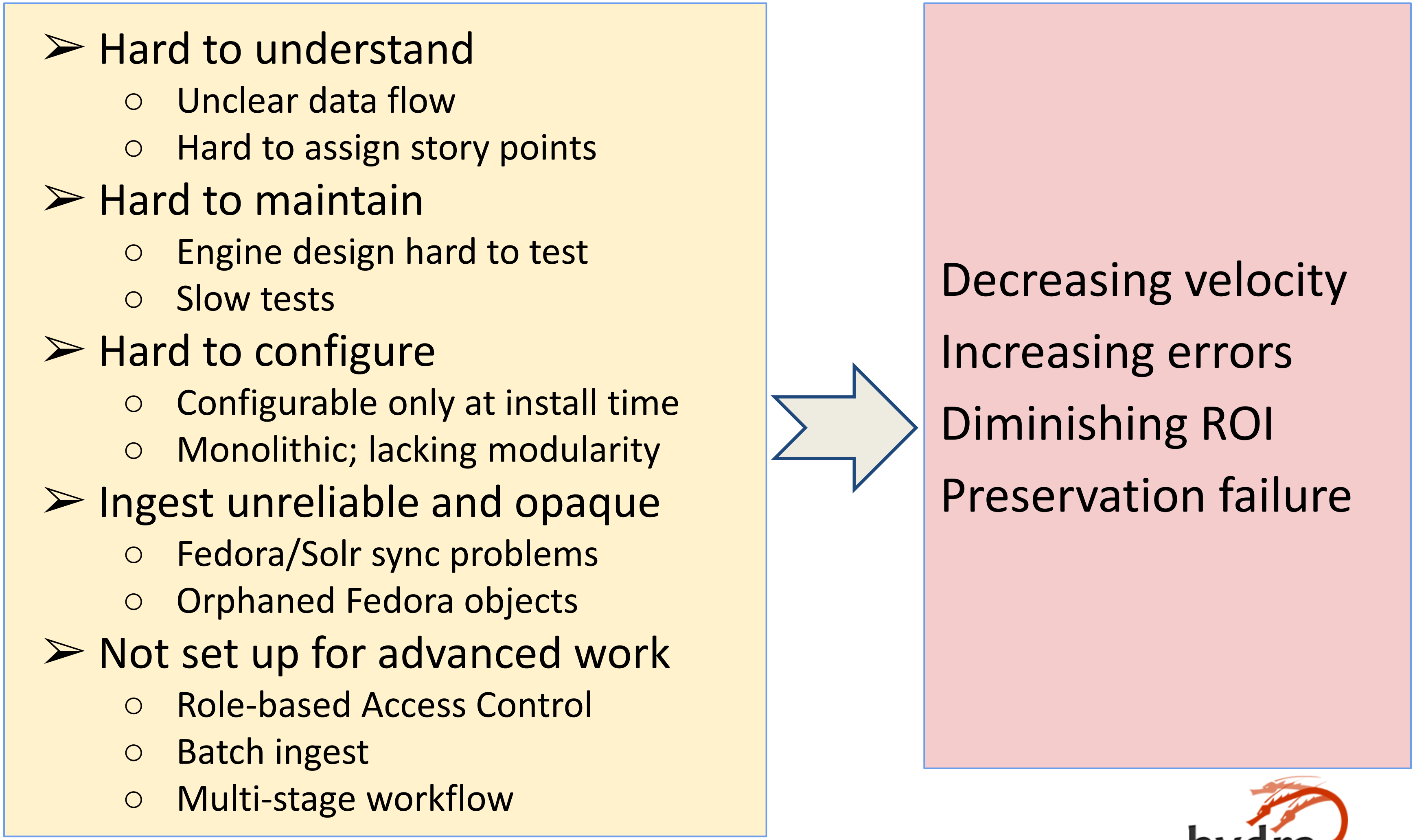
Hydramata

A set of separate hydra-based automata, "hydramata", that can be combined to form a single Hydra Head. Each hydramaton can be mixed and matched with other hydramata to provide wide format support tailored to fit your institutional or digital repository needs.

Transformation



Problems To Be Addressed



- Hard to understand
  - Unclear data flow
  - Hard to assign story points
- Hard to maintain
  - Engine design hard to test
  - Slow tests
- Hard to configure
  - Configurable only at install time
  - Monolithic; lacking modularity
- Ingest unreliable and opaque
  - Fedora/Solr sync problems
  - Orphaned Fedora objects
- Not set up for advanced work
  - Role-based Access Control
  - Batch ingest
  - Multi-stage workflow

Decreasing velocity

Increasing errors

Diminishing ROI

Preservation failure

Design Approach

- Separation of concerns into functional units:
  - Determine their responsibilities
  - Define their interfaces
  - Minimize coupling between functional units
- Tie units into a conceptual framework
  - Build out from ingest processing at the core
  - Treat works as arbitrary sets of metadata/content
  - Support interactive and batch processing equally
  - **Implement an Asynchronous transaction model**

Design Approach – Core IR

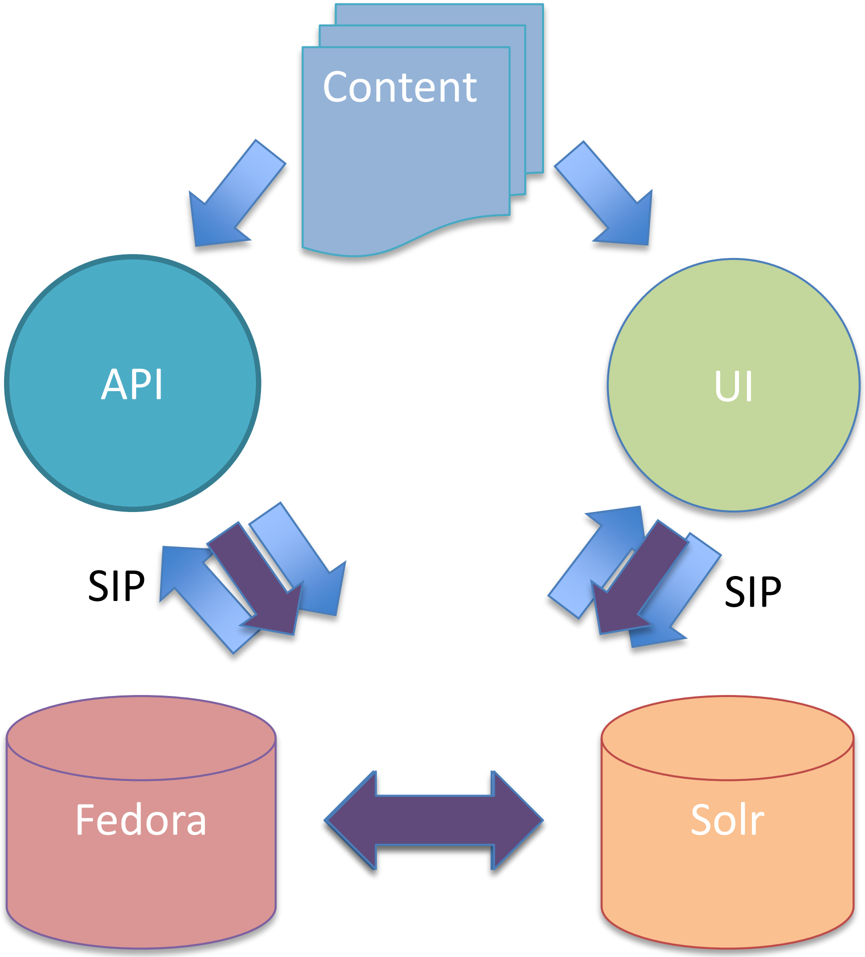
- ✓ Users can login to the system using their university credentials
- ✓ We are confident about the accuracy, persistence, security, and recoverability of what users upload
- ✓ Users can upload works via UI or batch
- ✓ Users can assign a license agreement to their works
- ✓ Works can be described, organized, discovered, viewed
- ✓ Works can be shared with (or restricted from) the right people
- ✓ Users can collaborate with others with individual works
- ✓ Users will get metrics about how their works are viewed
- ✓ There will be a variety of roles for people to interact with the IR
- ✓ Contributors names will link to standard identifiers (e.g. ORCID)
- ✓ DOIs can be assigned by users to works

Beyond Core IR

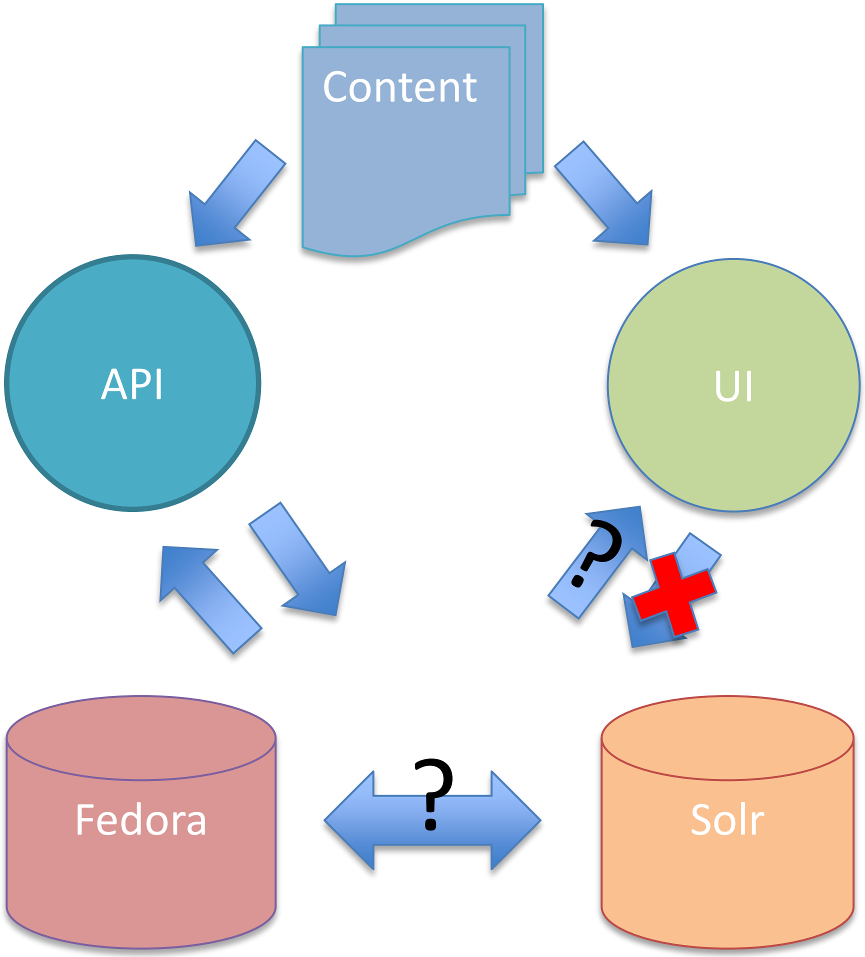
Components considered outside of the core:

- IIIF
- TEI support
- Specialized viewers
- ORCID plugin (minting ORCIDs from Hydramata)
- Minting DOIs from Hydramata
- Syndication to external systems (using OAI, ResourceSync or similar): SHARE, ORCID, Google Scholar, DPN, APTrust, DPLA
- Exhibits (Spotlight)
- Contributor profiles, links to campus research information systems (VIVO)

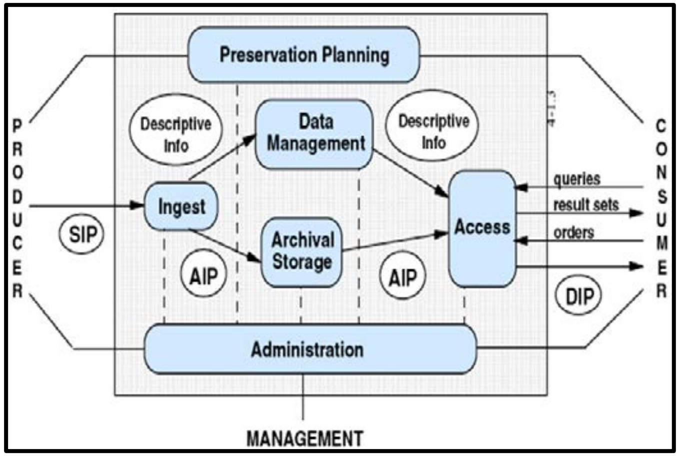
Single Asynchronous(?) Transaction



Single Asynchronous(?) Transaction



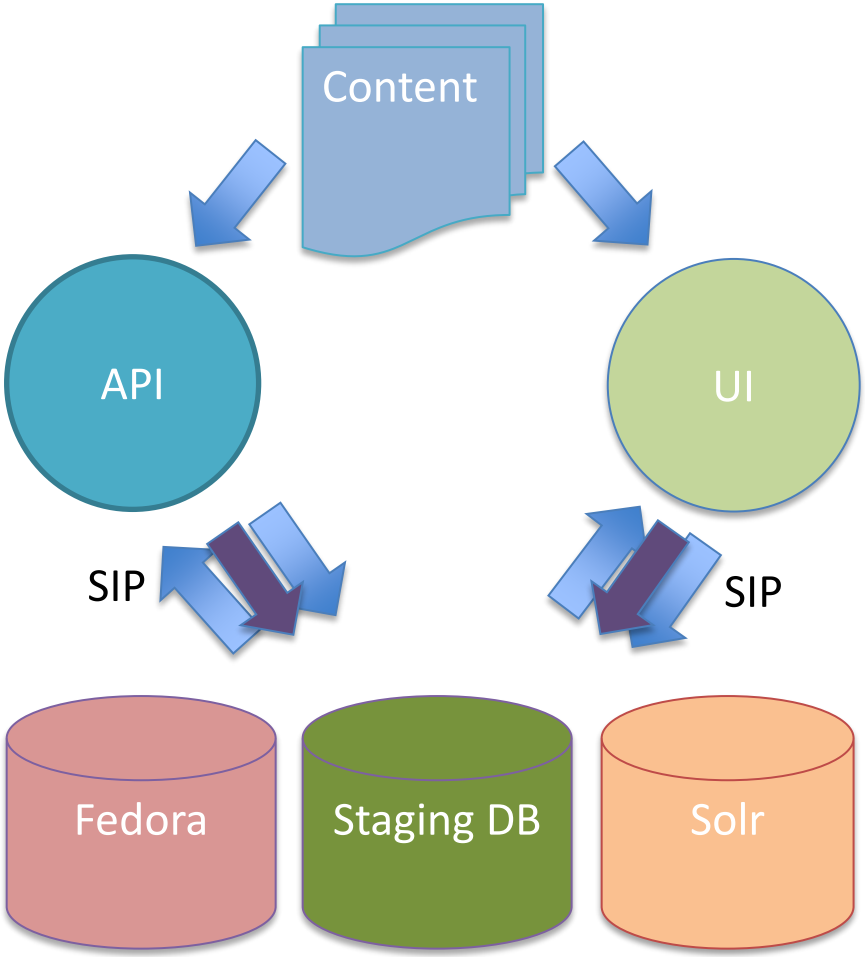
Ingest Processing



Build on OAIS Reference Model concepts:

- SIP - Submission Information Package
  - Gather user input into a SIP until validated
  - Maintain SIP until validated
- AIP - Archival Information Package
  - Package work into Fedora Object(s)
  - Manage storage into Fedora
- DIP - Dissemination Information Package
  - Aggregate work content and metadata in a common form
  - Provide in-progress and submitted works consistently

Single Asynchronous(~~?~~) Transaction

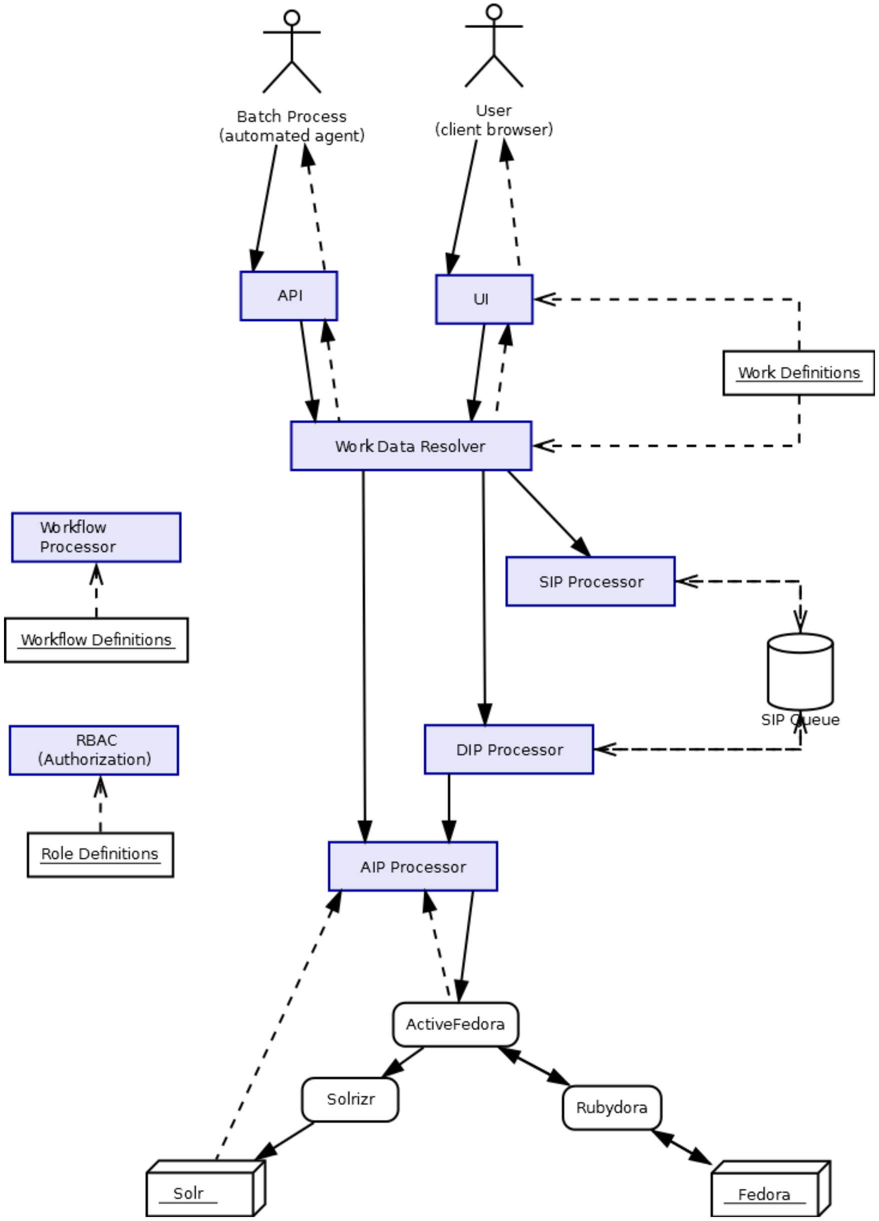


Rebuild the Core

Yet to determine where some elements fit in:

Role-Based Access Control

Multi-stage workflow support



hydra

Problems Addressed

- Easier to understand
  - Design model clarifies overall function
  - Modularity makes tasking less difficult
- Easier to maintain
  - Created as an app not an engine
  - Decoupling helps test creation/execution
- Easier to configure
  - More intelligence in the core
  - Variation in works encapsulated in data
- Ingest process built to handle errors
  - Fedora/Solr updated only with valid data
  - No orphaned Fedora objects
- Accommodates advanced work
  - Reference implementations of modules allow programming to the interface while deferring full implementations

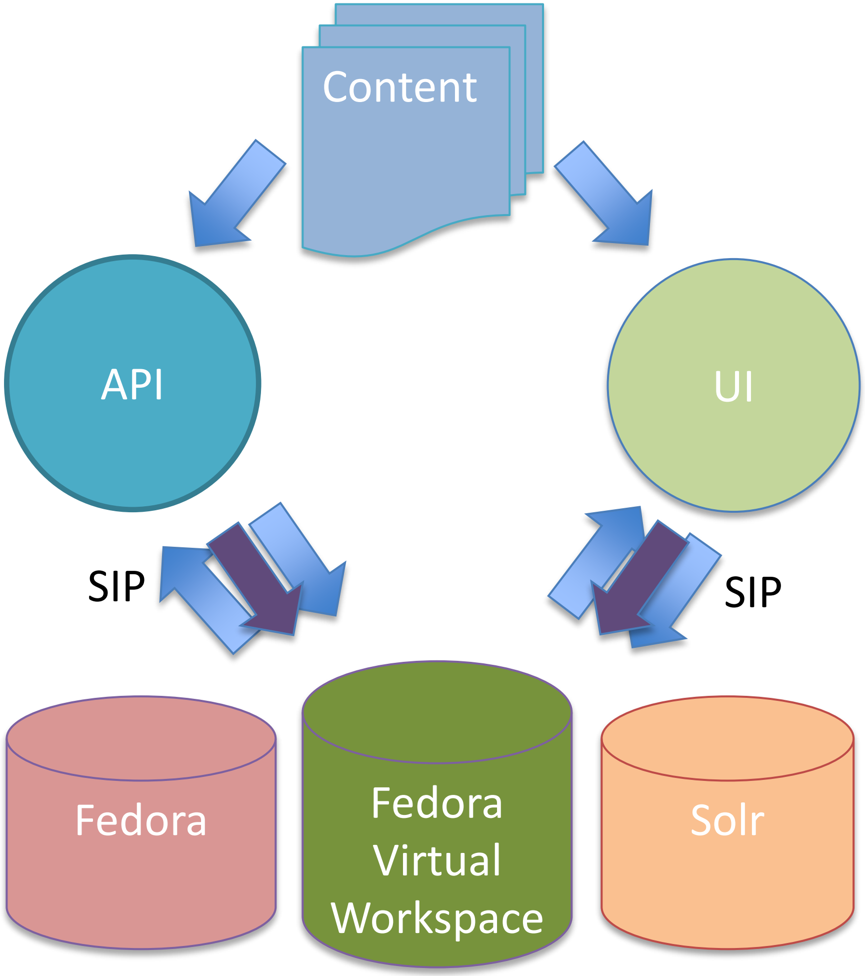
Steady velocity

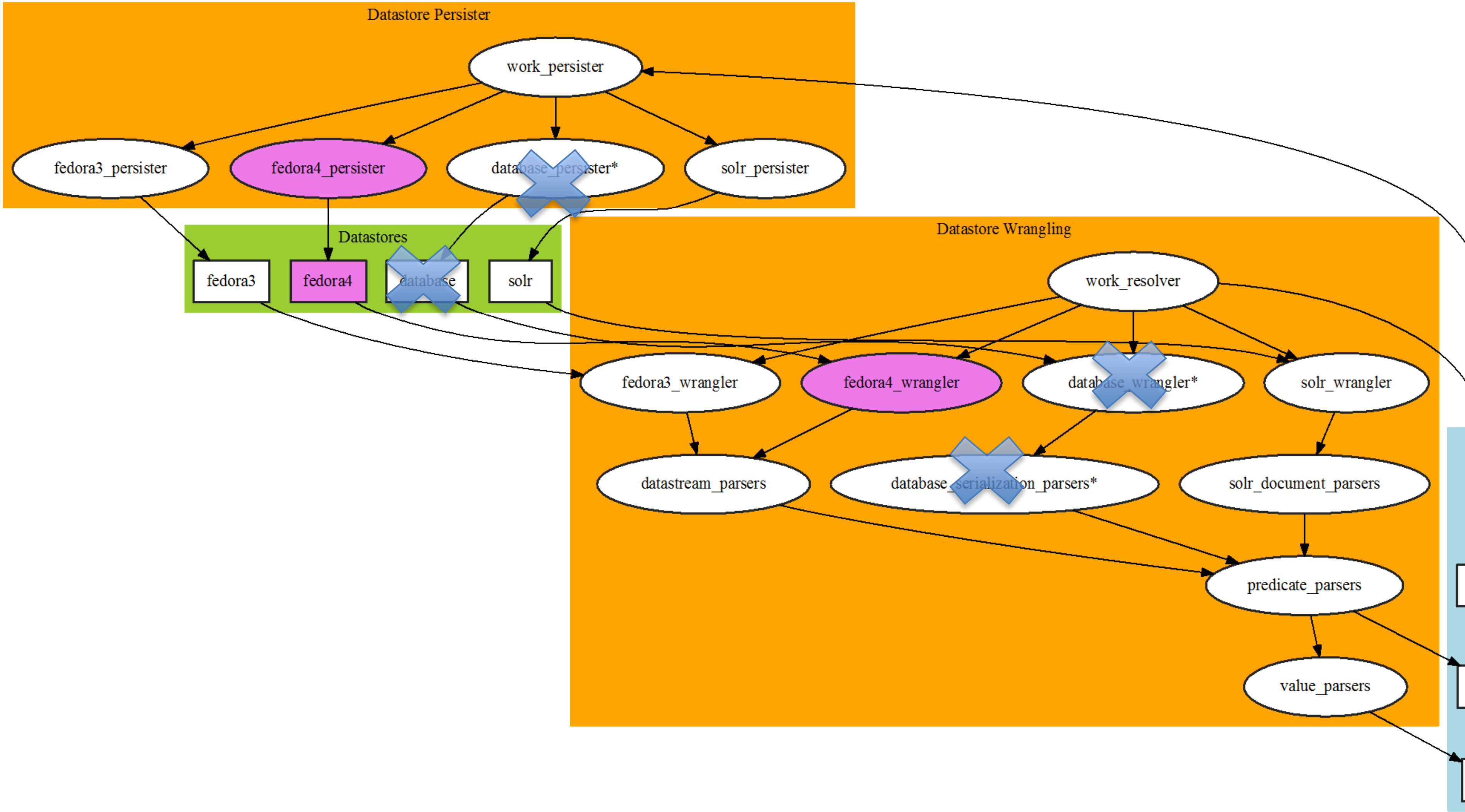
Decreased errors

Increasing ROI

Preservation

What About Fedora 4?



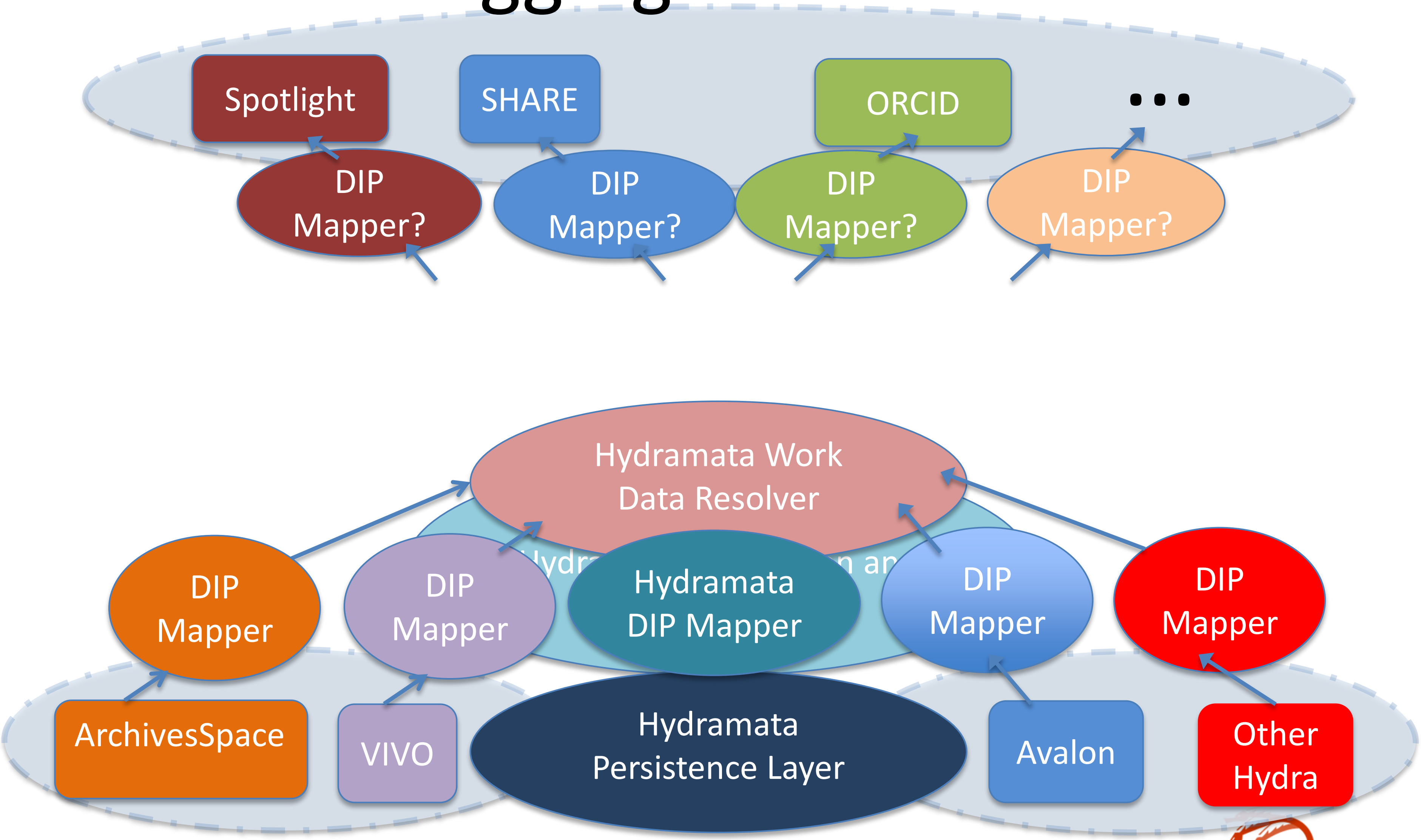


Hydramata::Work Architecture

Swappable Hydramata

	Images	Datasets	Articles	ETDs
Deposit	Image Metadata Fields	Dataset Metadata Fields	Article Metadata Fields	ETD Metadata Fields
Mediate	Image Approval Workflow	Dataset Approval Workflow	Article Approval Workflow	ETD Approval Workflow
Ingest	Image Processing Pipeline	Dataset Processing Pipeline	Document Processing Pipeline	
Discovery	Image Gallery Views	Default Search Behavior	Default Search Behavior	Default Search Behavior
Viewer	Zoomable Image Viewer	Dataset Visualization Tools	Paginated Document Viewer	
Disseminator	Tiling Image Server	Default File Disseminator	Default File Disseminator	Default File Disseminator

Plugging In and Out



Spotlight

SHARE

ORCID

• • •

DIP Mapper?

DIP Mapper?

DIP Mapper?

DIP Mapper?

Hydramata Work Data Resolver

DIP Mapper

DIP Mapper

Hydramata DIP Mapper

DIP Mapper

DIP Mapper

ArchivesSpace

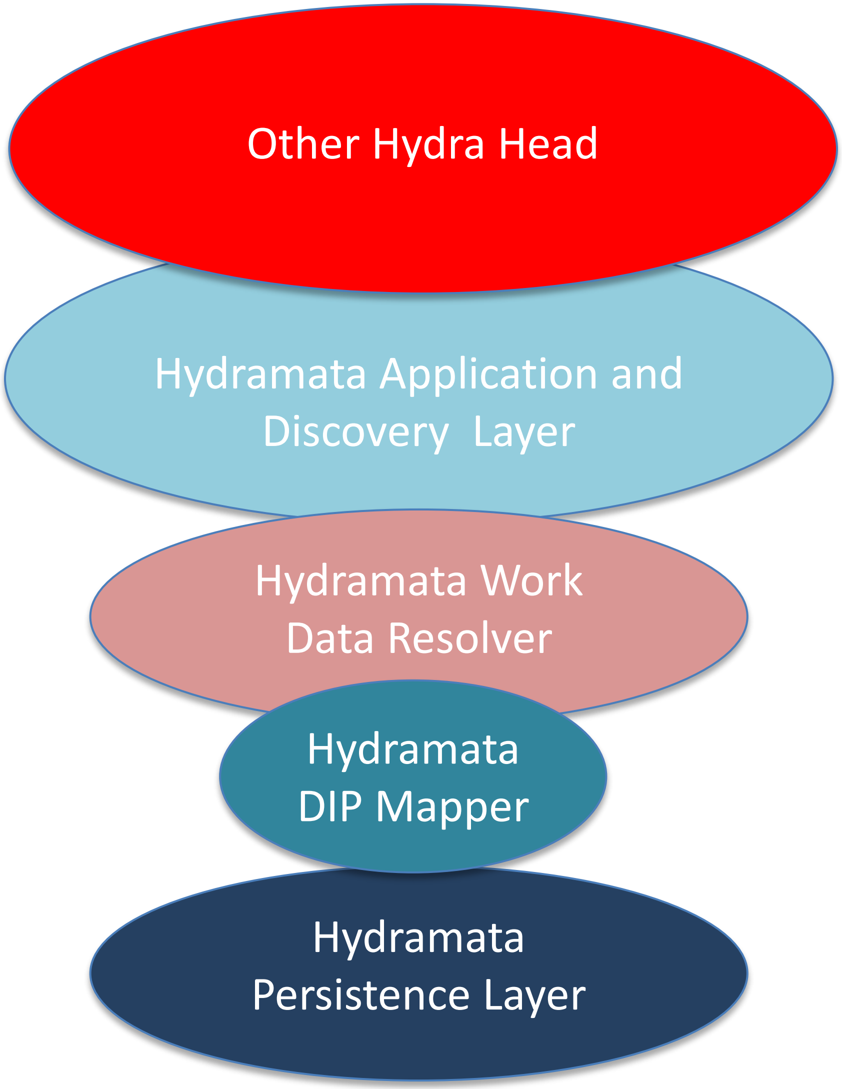
VIVO

Hydramata Persistence Layer

Avalon

Other Hydra

Plugging In and Out



Other Hydra Head

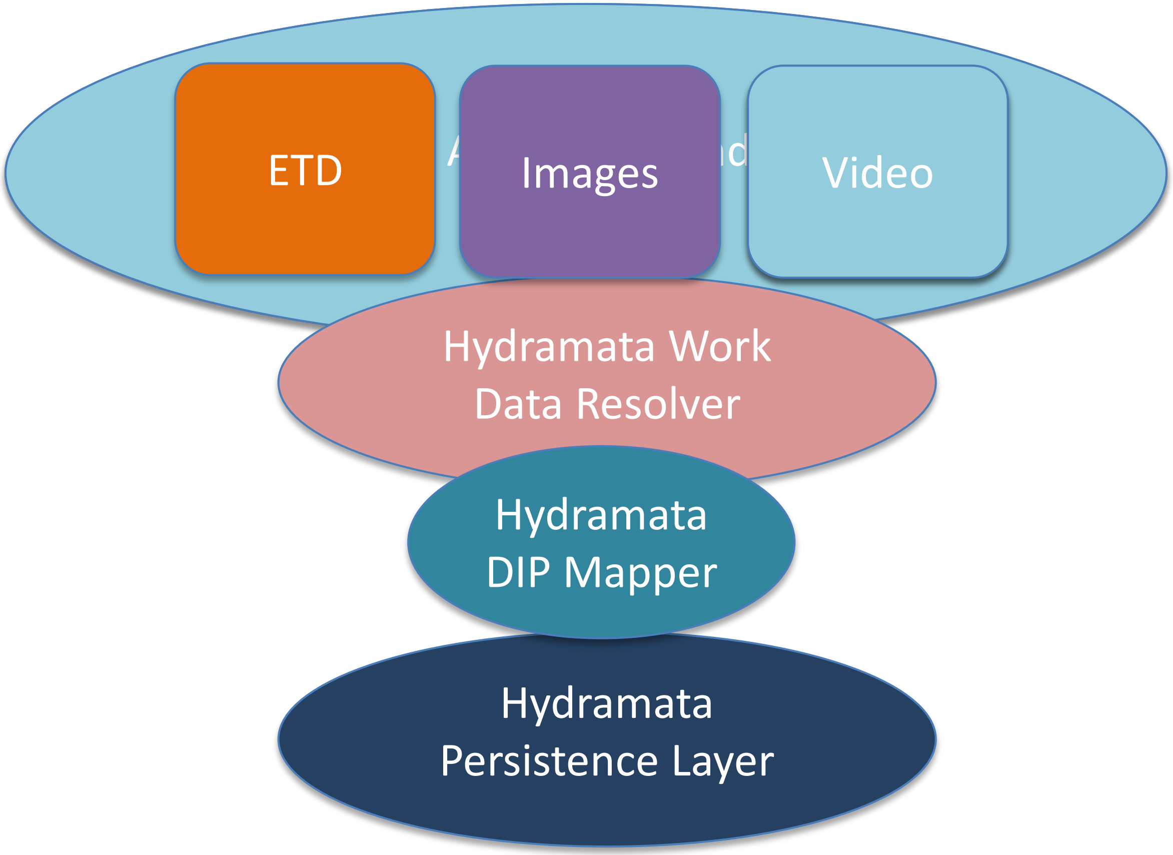
Hydramata Application and Discovery Layer

Hydramata Work Data Resolver

Hydramata DIP Mapper

Hydramata Persistence Layer

Plugging In and Out



Acknowledgements

- Jeremy Friesen, Technical Lead
- Dan Brubaker Horst, Product Owner
- Paul Clough, Developer
- Ray Lubinsky. Product Owner
- Sue Richeson, Developer
- Julie Rudder, Lead Product Owner
- Mike Stroming, Developer
- Entire Hydramata Project Team (15+ individuals)

Contact Info

- Rick Johnson, rick.johnson@nd.edu
- Linda Newman, newmanld@ucmail.uc.edu
- Project Hydramata wiki:
  - https://wiki.duraspace.org/display/hydra/Hydramata+Project
- Curate code base:
  - https://github.com/projecthydra-labs/curate
- Hydra Community:
  - http://projecthydra.org/

Q & A

http://en.wikipedia.org/wiki/Helsinki_Workers%27_House#mediaviewer/File:Paasitorni_by_the_sea.jpg
http://creativecommons.org/licenses/by-sa/3.0/