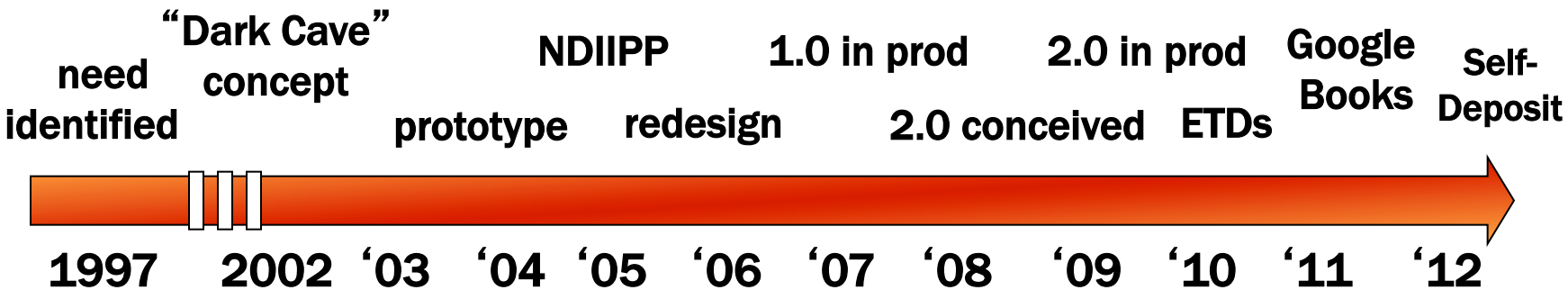
STANFORD UNIVERSITY LIBRARIES

The Stanford Digital Repository (SDR), Fedora & Hydra

SDR is…

- a core Library service
- Stanford's digital preservation system
- an increasingly robust digital asset management & access system
- in production since Dec 2006



http://library.stanford.edu/sdr

SDR Preservation Core

The Stanford Digital Repository (SDR) provides services to make scholarly resources

available over the long term

by helping ensure their

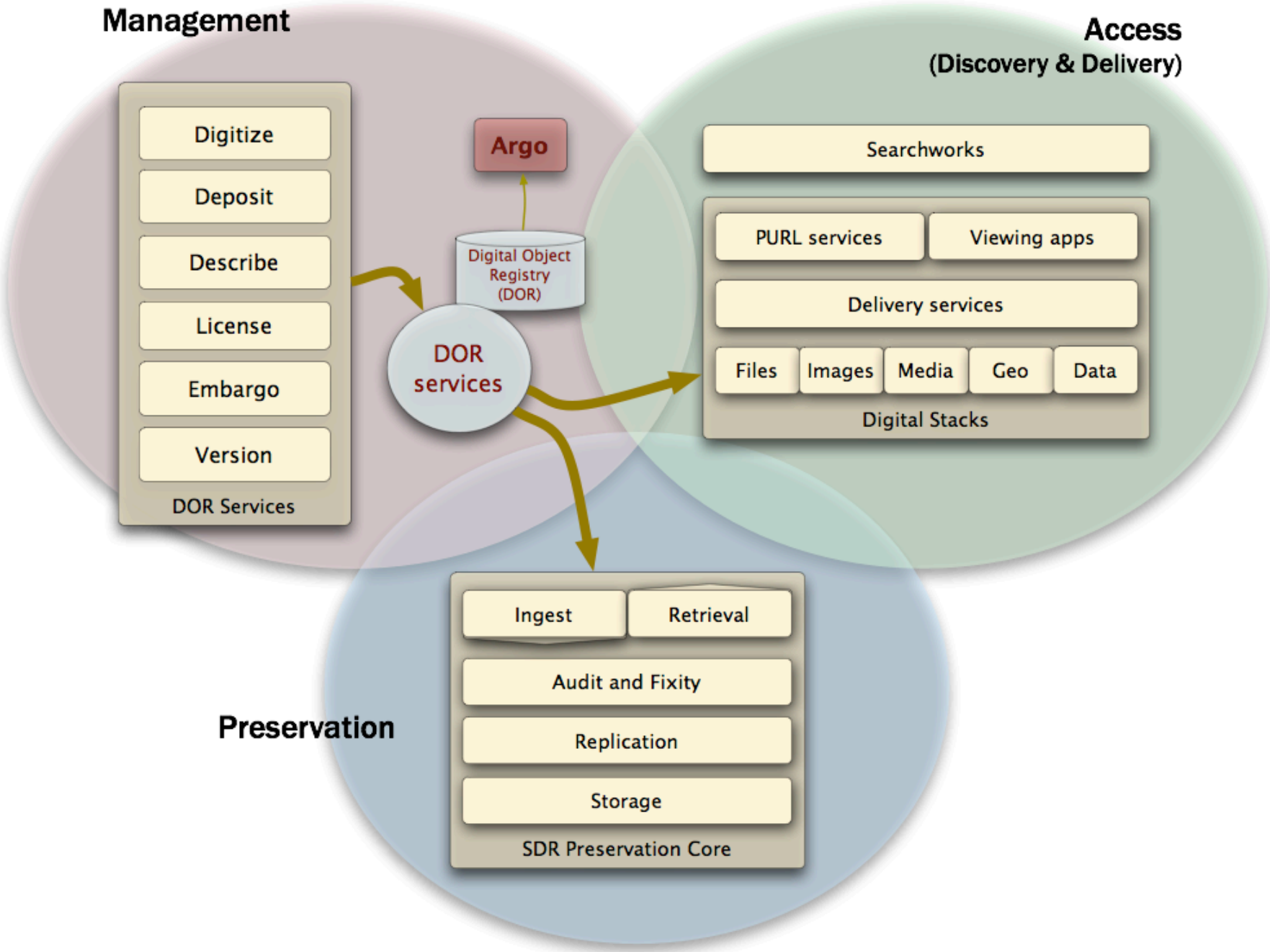
integrity,

authenticity, and

reusability.

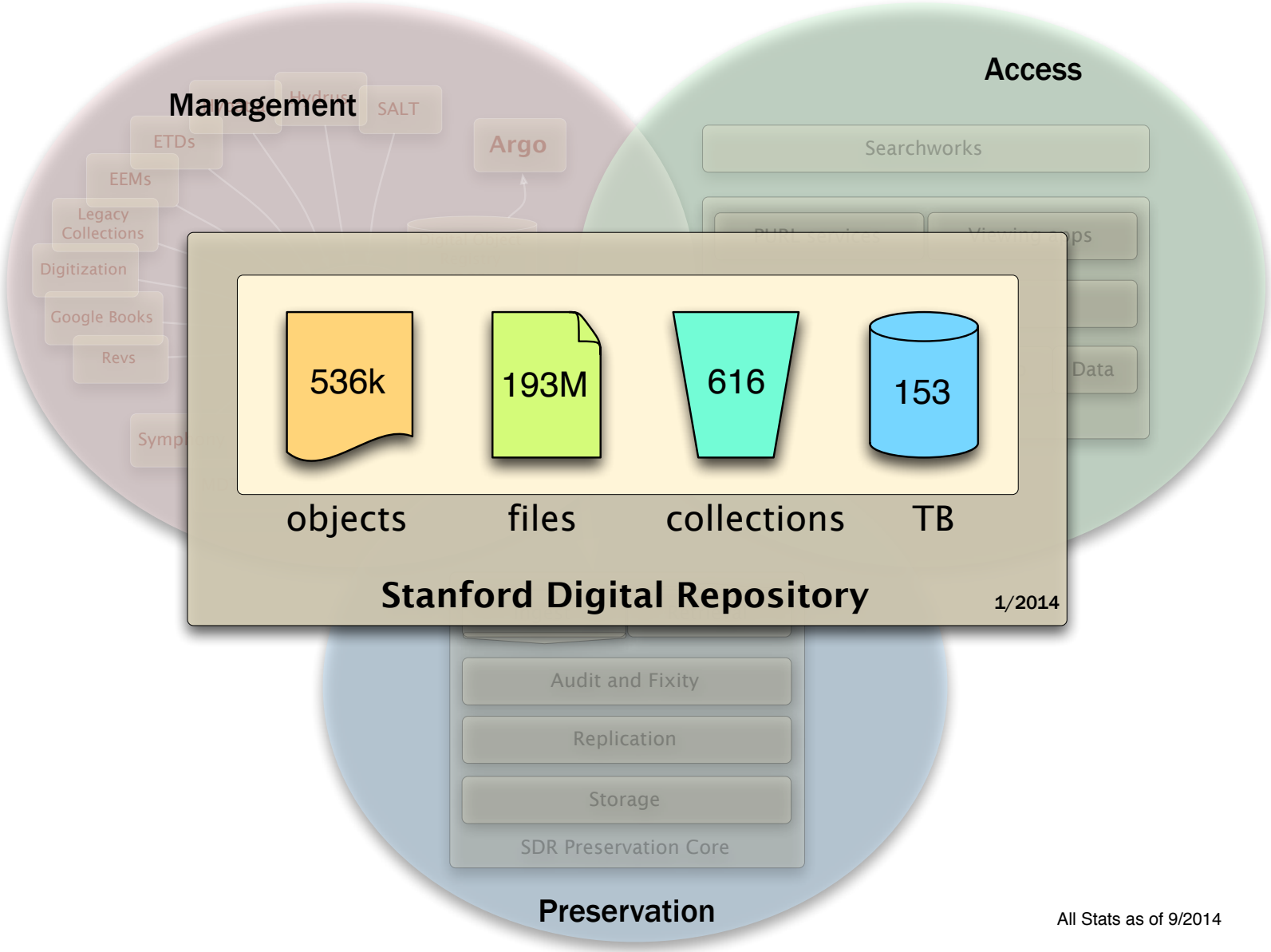
To fulfill its mission, the SDR must be

secure, sustainable and trustworthy.


Management
Access
(Discovery & Delivery)
Digitize
Deposit
Describe
License
Embargo
Version
DOR Services
Argo
Digital Object Registry (DOR)
DOR services
Searchworks
PURL services
Viewing apps
Delivery services
Files
Images
Media
Geo
Data
Digital Stacks
Ingest
Retrieval
Audit and Fixity
Replication
Storage
SDR Preservation Core
Preservation

The Stanford Digital Repository

Contents as of Sept 2014

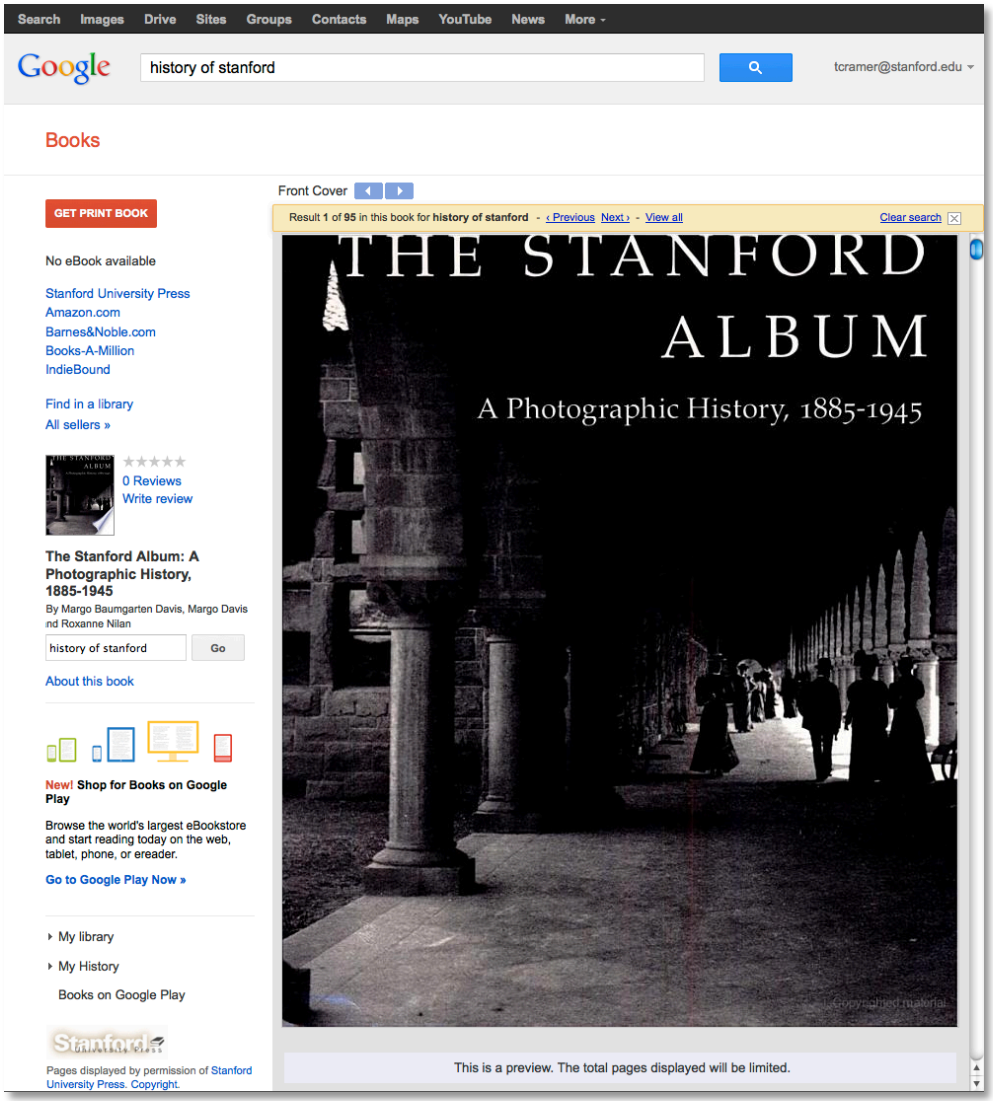


All Stats as of 9/2014

E.g., Google-Scanned Books

Download, process and preserve scanned volumes in SDR for...

- local indexing,
- text mining,
- selective delivery, and
- long-term access.



E.g., Monterey Jazz Festival

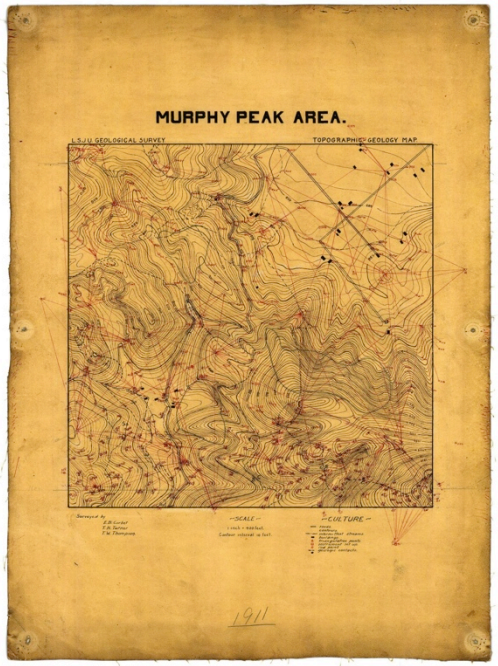
- Festival founded in 1958: longest running jazz festival in the world.
- Rich collection of recordings from inception, spanning over 50 years, in varying states of condition & decay.
- Archives held at Stanford's Archive of Recorded Sound
- ~800 audio recordings, 1.6 TB audio files in SDR
- ~250 video recordings, 22 TB video files in SDR

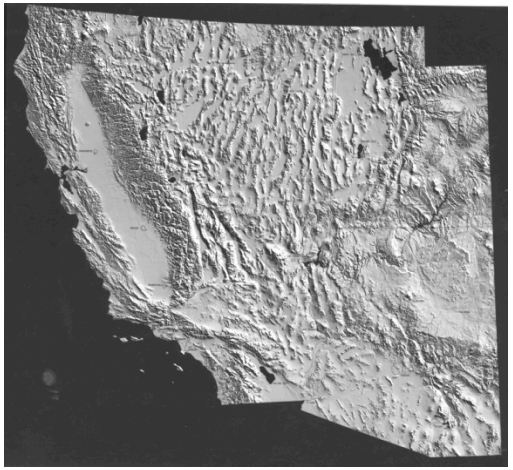
Access:

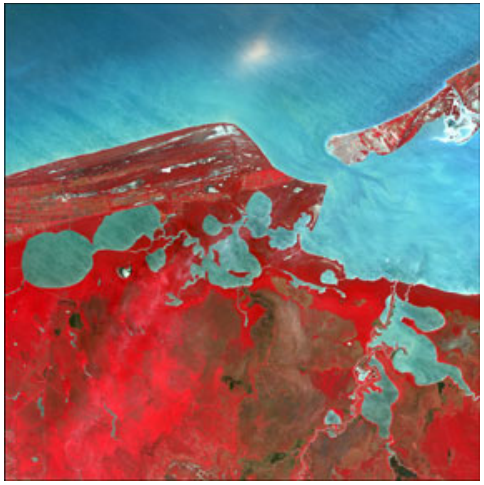
- complete database of digital recordings online at *collections.stanford.edu/mjf*
- Access via in-site visit to ARS
- New commercial releases on MJF Records

E.g., National Geospatial Digital Archive

- Some 27,000 “at risk” geospatial objects
- TIFFs, GeoTIFFs, Shapefiles, Digital Elevation Models, Digital Orthophoto Quadrangle files





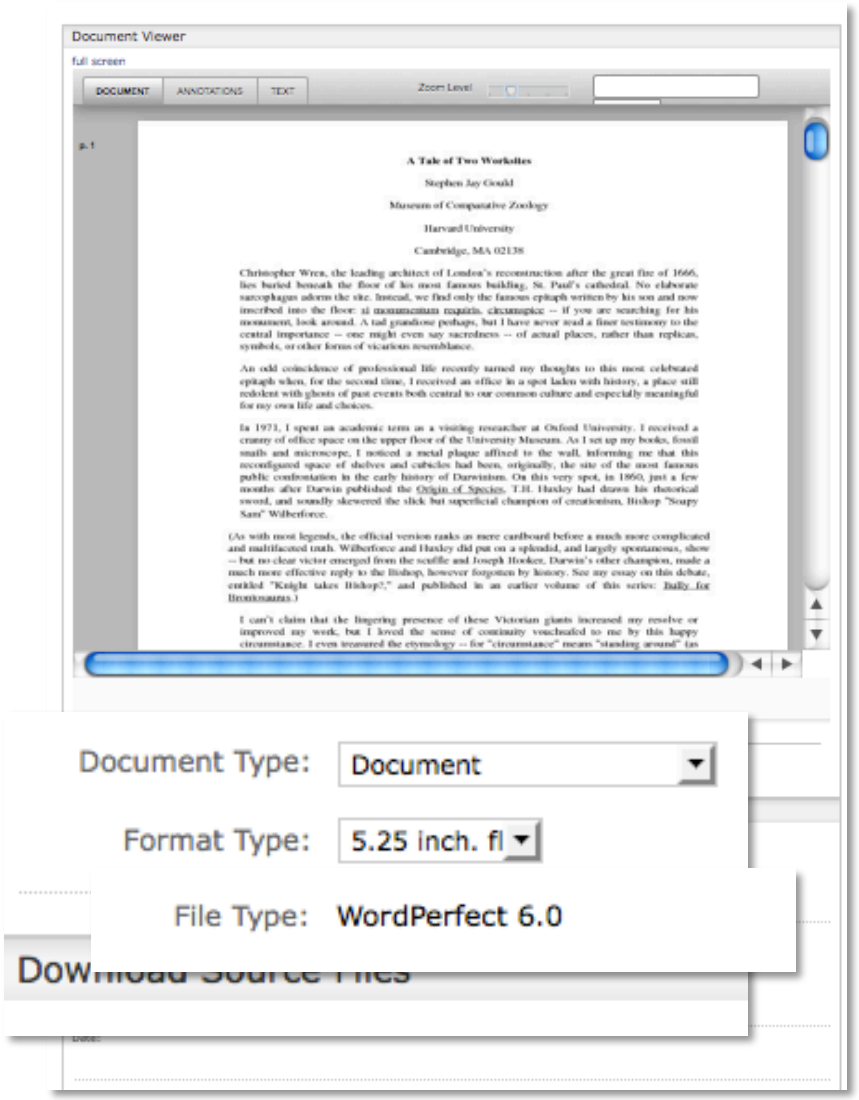


E.g., Preserving Virtual Worlds

Stanford University Libraries
Second Life Open House,
31 July 2009

E.g., Forensically Extracted Born Digital Files

- Digital Forensics lab extracting original computer files from legacy media
- Actively building pipeline from extraction to preservation store
- Support for both immediate and deferred archival processing & description



Technology

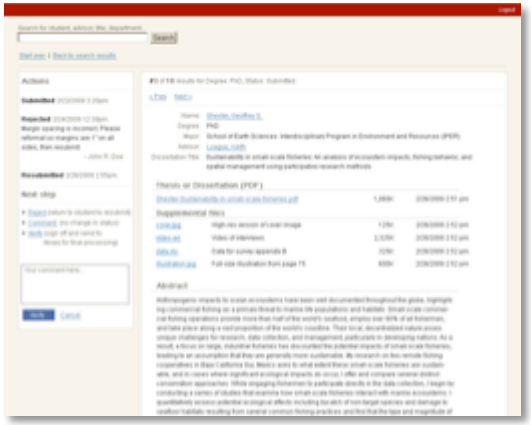
Fedora™

hydra

IBM Enterprise Tape Library

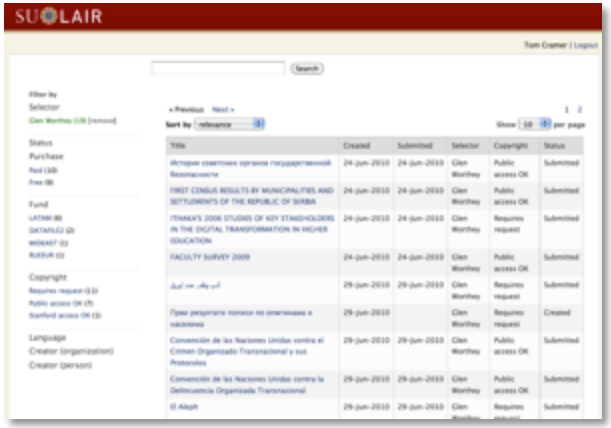
NetApp 6080 Disk

Hydra-based Applications at Stanford



ETD's – Electronic Theses & Dissertations

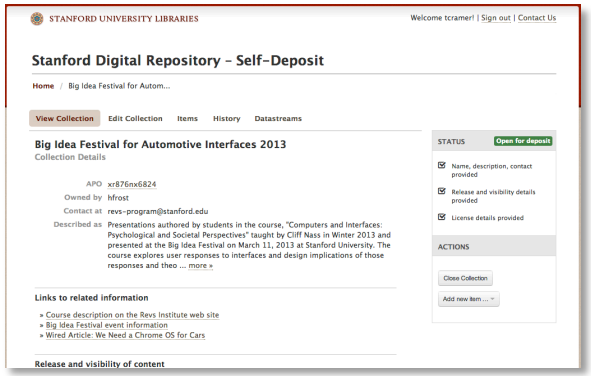
SALT – Self-Archiving Legacy Toolkit



EEMs – Everyday Electronic Materials

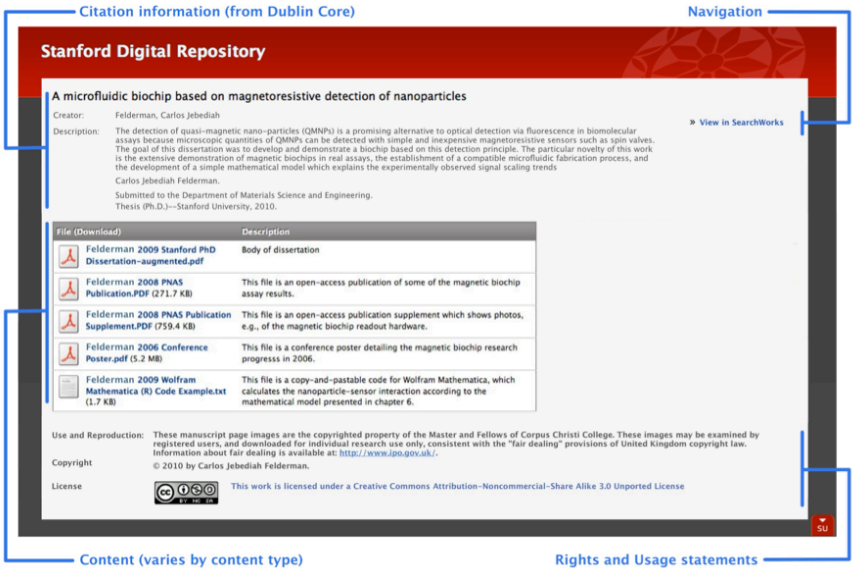
Argo – Repository Reporting and Management

Hypatia – Archives & Special Collections



SDR – Web UI (*aka Hydrus*)

Access Environments



PURL (Persistent URLs) Landing Pages

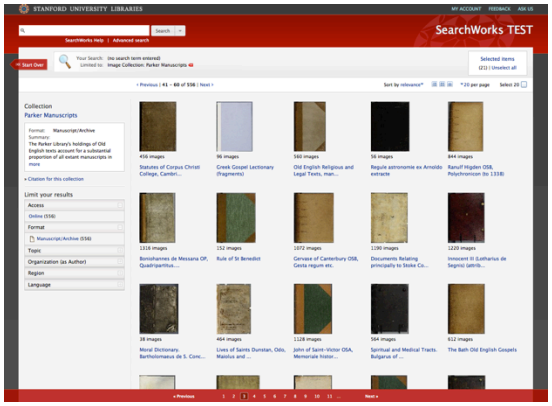
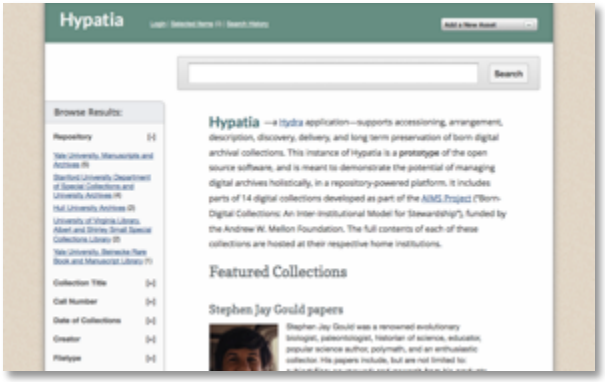


Image Galleries

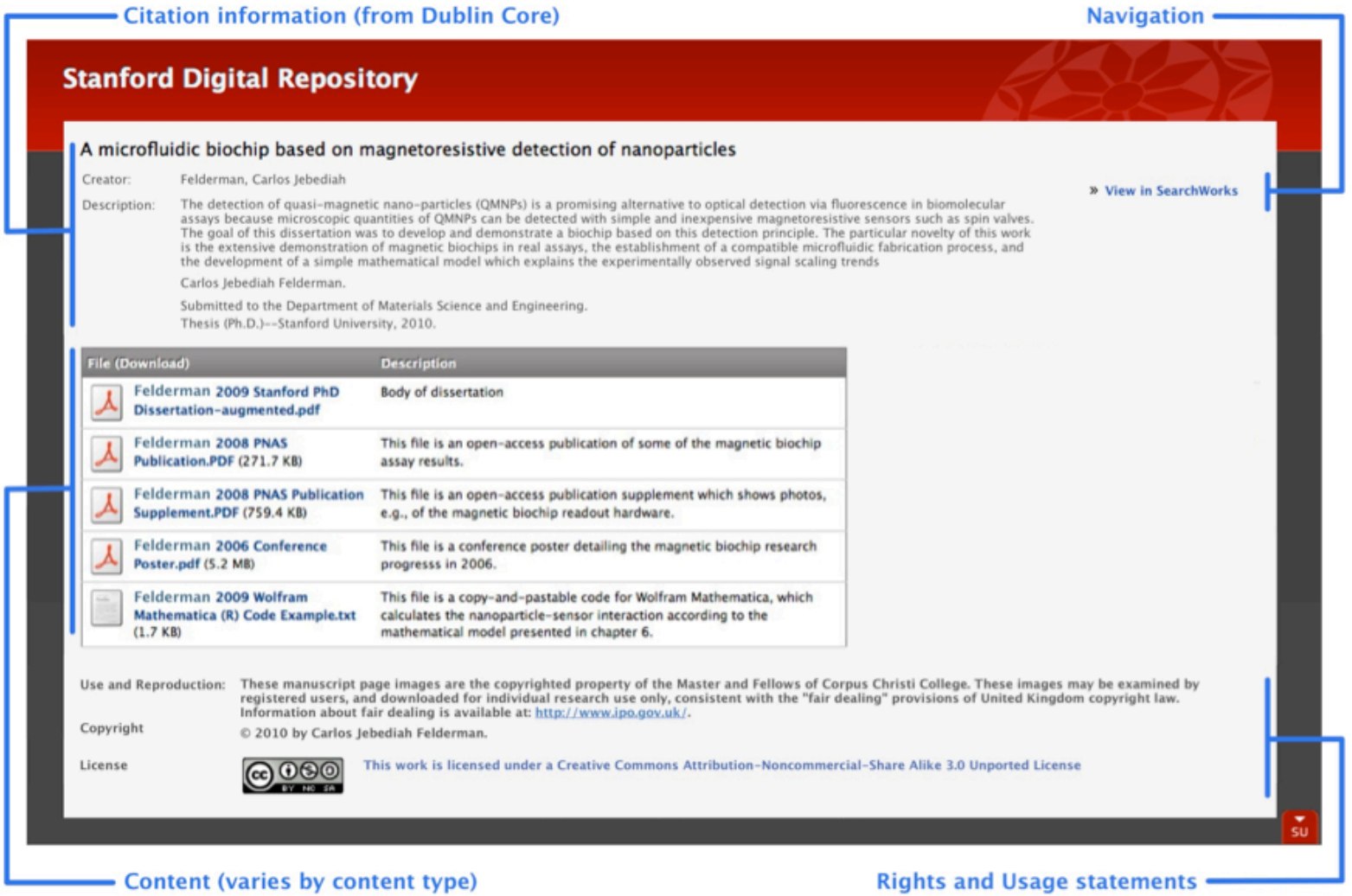


Embeddable Web Widgets

International Image Interoperability Framework

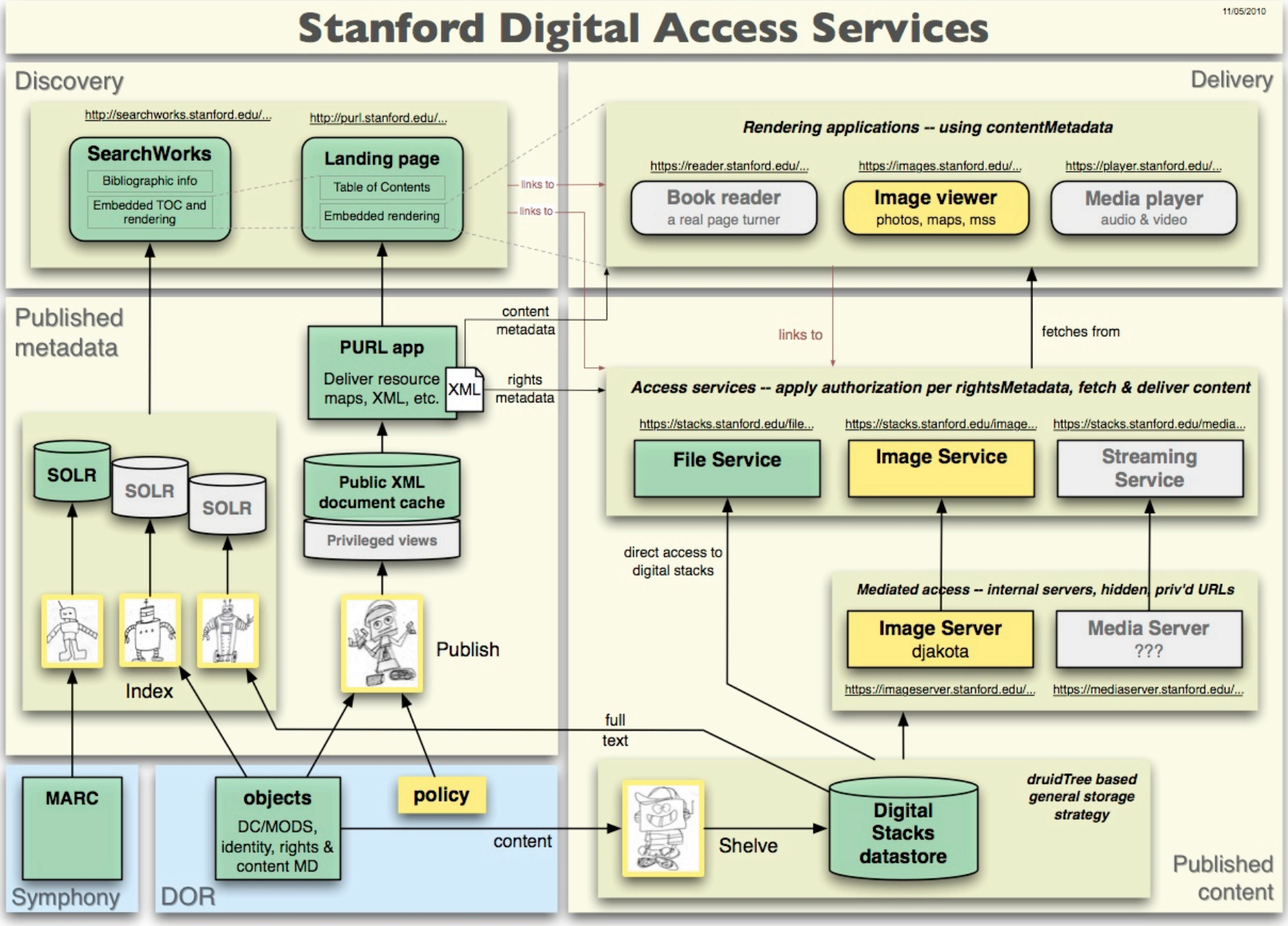
http://iiif.io

PURLs (Persistent URLs)



Stanford Digital Access Services

11/05/2010



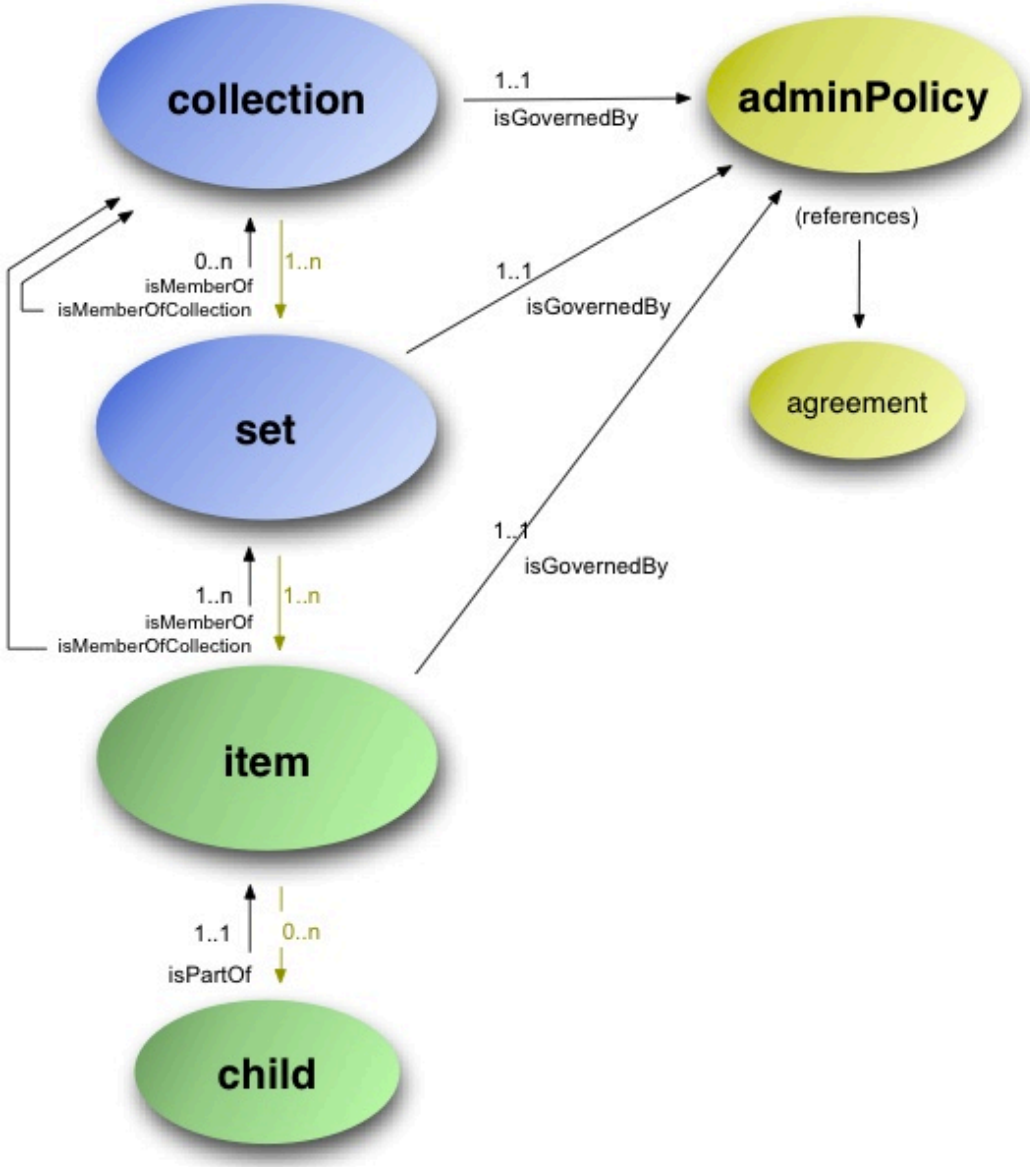
Standard Data Streams

DOR datastream	Source	Description	Fedora Control Group
identityMetadata	Stanford	Core identifying information: title, creator, source and other identifiers, tags, etc.	Internal XML
descMetadata	Hydra	Currently MODS xml	Managed Content
technicalMetadata	Hydra	JHOVE generated metadata describing the content files that make up the digital object.	Managed Content
sourceMetadata	Hydra	Information identifying the source of the scanned object	Managed Content
rightsMetadata	Hydra	Holds human and machine readable expressions of Access and Use permissions/restrictions.	Managed Content
provenanceMetadata	Hydra	Contains a block of provenance information for each agent that handled the record.	Managed Content
contentMetadata	Hydra	Information about object structure, file sequence and type, and checksums.	Managed Content
RELS-EXT	Fedora	RDF triples describing selected object attributes and inter-object relationships.	Internal XML

Standard Data Streams - continued

DOR datastream	Source	Description	Fedora Control Group
versionMetadata	Stanford	A summary history of versions that have been accessioned/archived	Internal XML
workflows	Stanford	The workflows datastream is a view into all the workflows defined for an object, current and past. It is a concatenation of individual workflow XML in a <workflows> document container.	Redirect Referenced Content
DC	Fedora	Dublin Core XML, reflects author, title and identifier information here to take advantage of Fedora indexing for their administrative tools only.	Internal XML
embargoMetadata	Stanford	Operational metadata for managing embargo settings, extensions and release	Internal XML
events	Stanford	A lightweight history of events relating to the management of the object, e.g., embargo extensions and release.	Internal XML
varies	Stanford	Applications may have their own datastream(s) to manage local settings and other pertinent information.	Internal XML

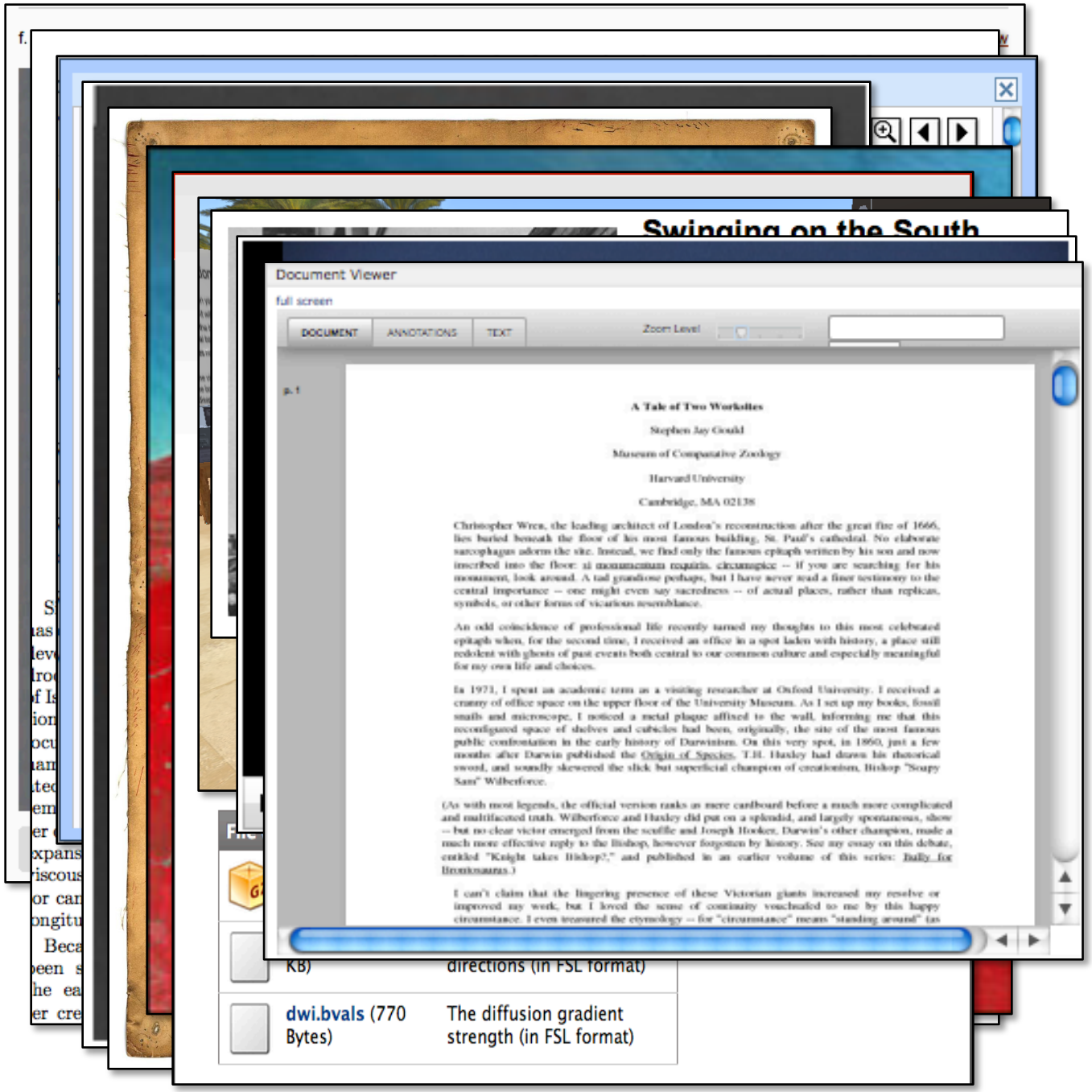
Basic Object Model



hydra

What Is Hydra?

- A robust repository fronted by feature-rich, tailored applications and workflows (“heads”)
  - ➪ *One body, many heads*
- Collaboratively built “solution bundles” that can be adapted and modified to suit local needs.
- A community of developers and adopters extending and enhancing the core
  - ➪ *If you want to go fast, go alone. If you want to go far, go together.*



Books

Articles

Theses

Images

Maps

Data (Raster)

Data (Comp.)

Data (Observ.)

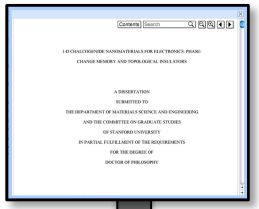
Audio

Video

Documents

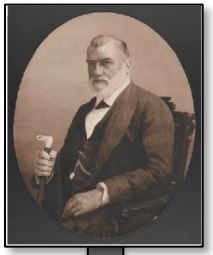
Point Solution Approach ...Welcome to Siloville

ETDs (Theses)

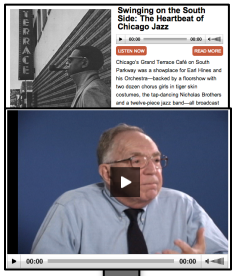


Books, Articles

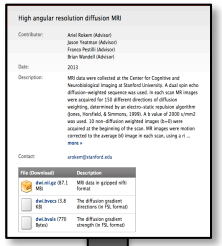
Images



Audio-Visual

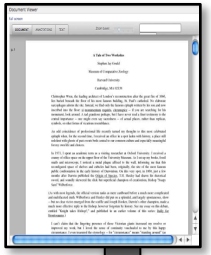


Research Data



Maps & GIS

Documents



ETD

IR

Image DB

DAM

?

Geospatial Inf.

Records Mgmt.

Management

Access

Preservation(?)

Effective?

Sustainable?

Repository-Powered Approach

ETDs (Theses)

Books, Articles

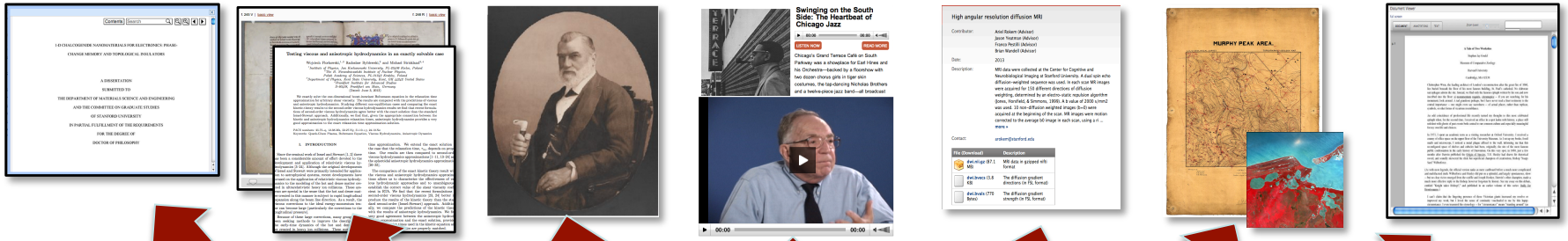
Images

Audio-Visual

Research Data

Maps & GIS

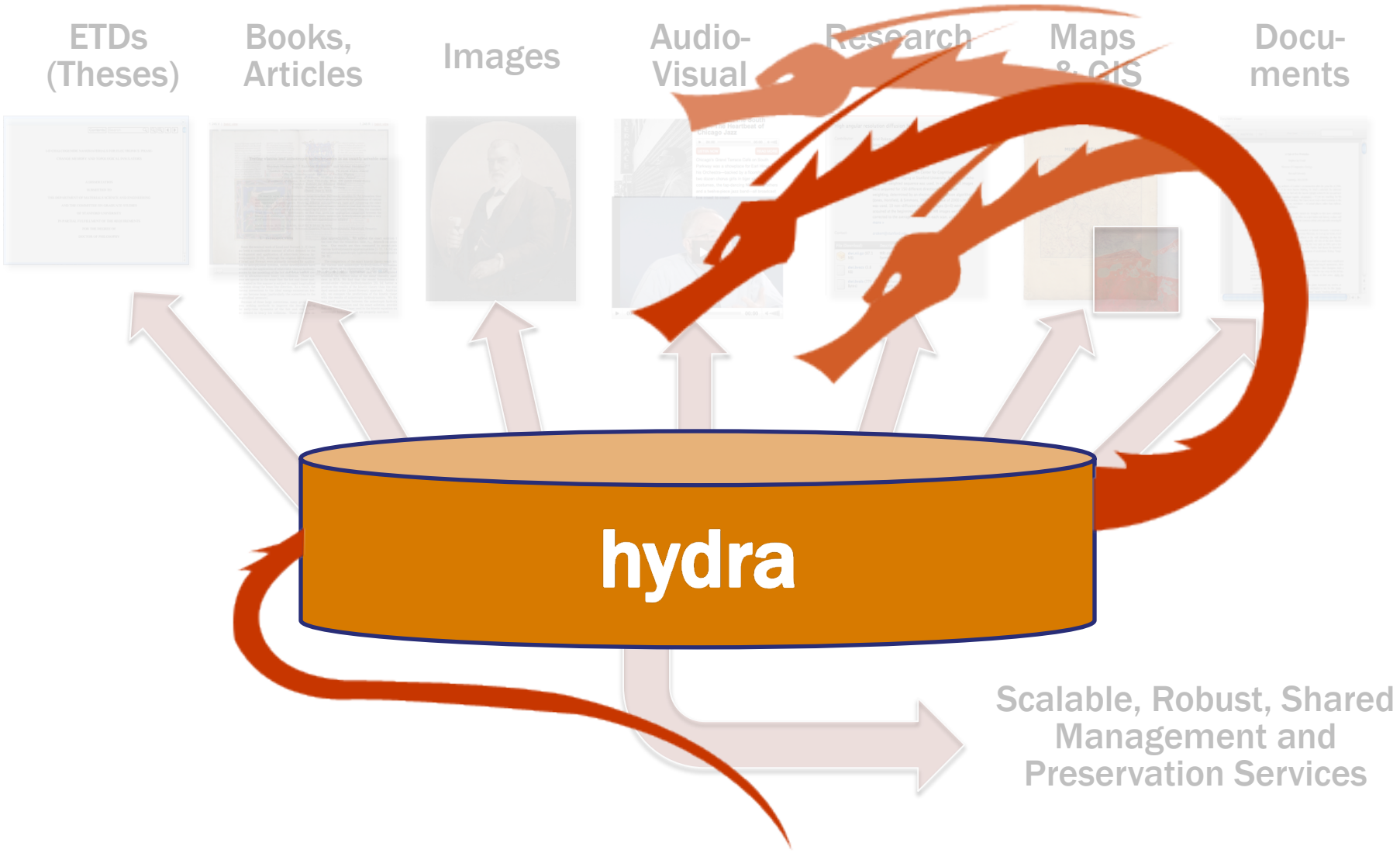
Docu-ments



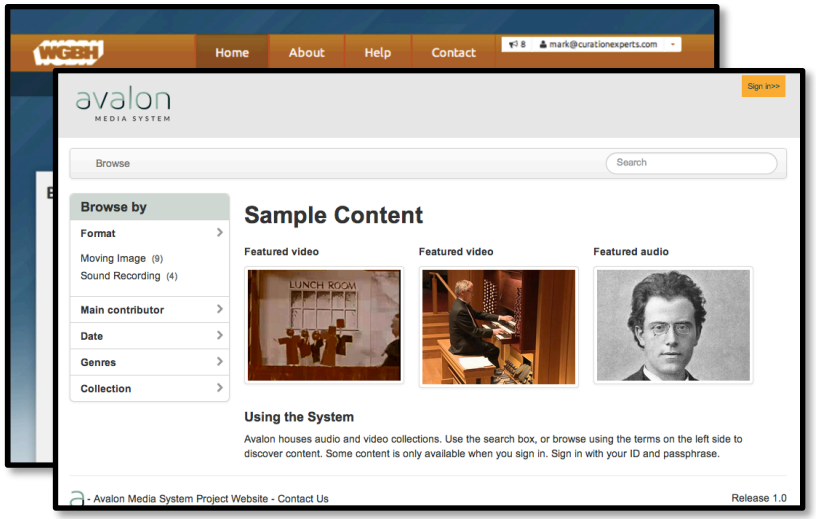
Digital Repository

Scalable, Robust, Shared Management and Preservation Services

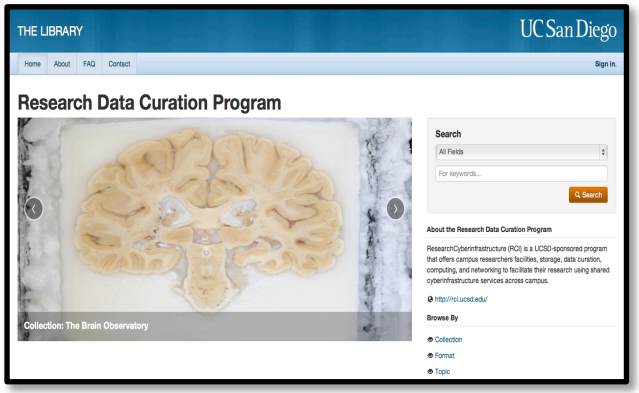
One Body, Many Heads...



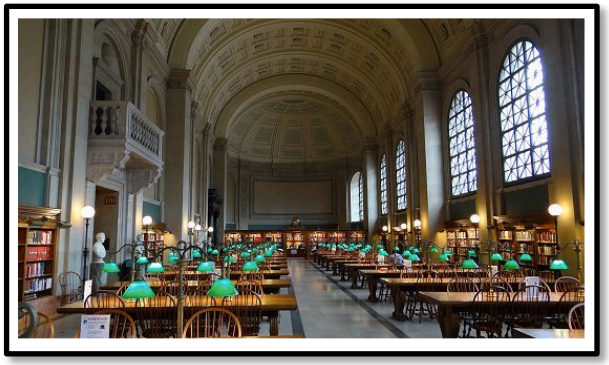
Hydra Heads of Note



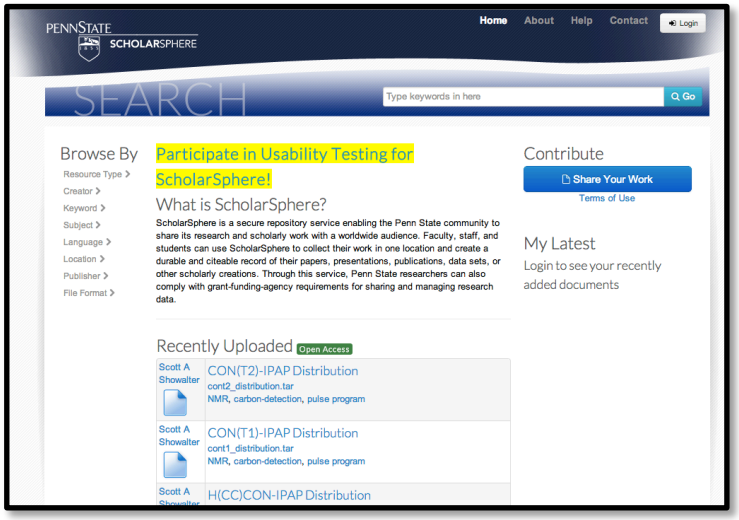
Avalon & HydraDAM for Media



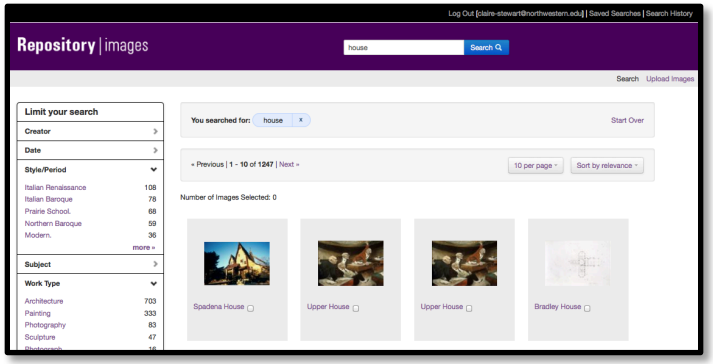
UCSD DAMS



BPL Digital Commonwealth

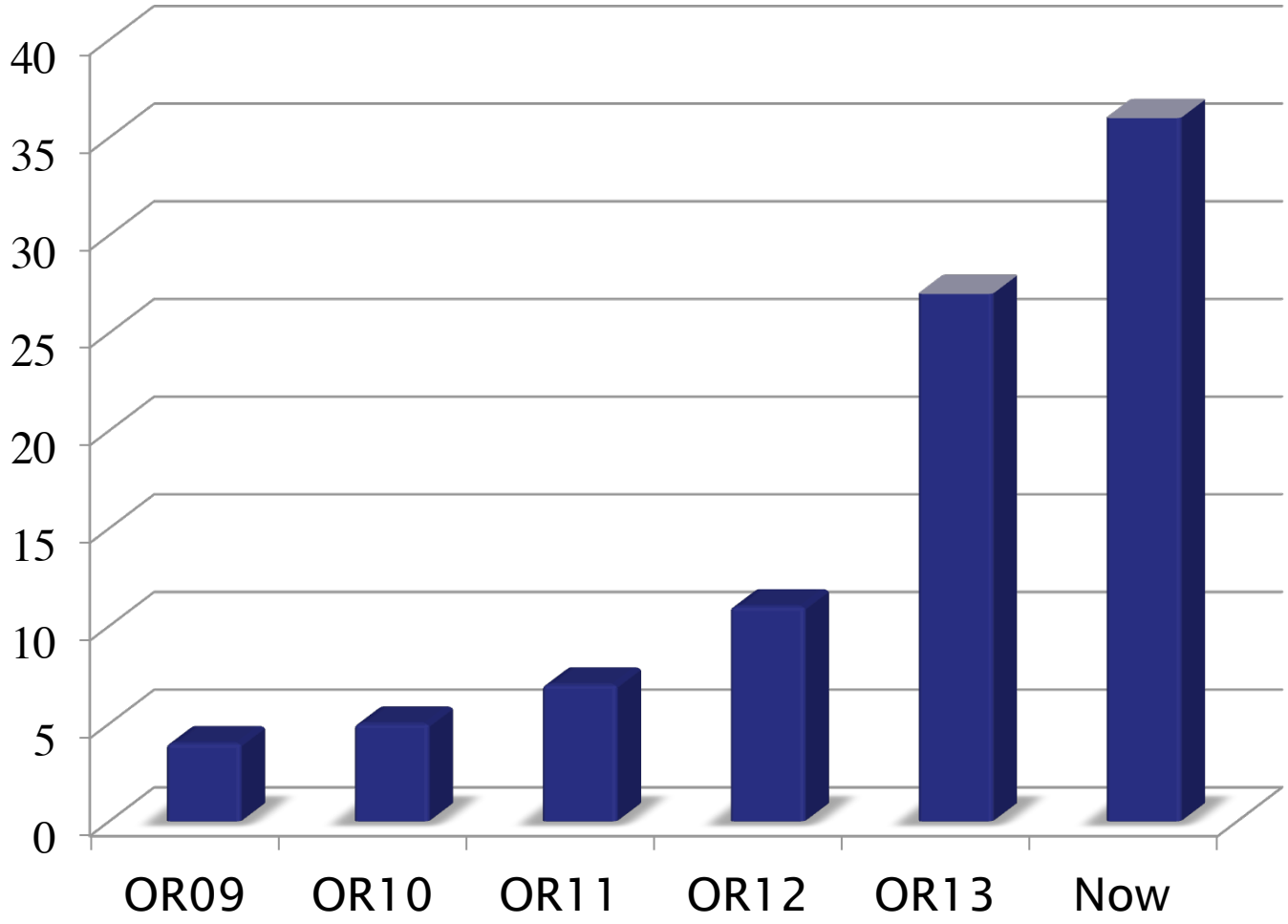


Sufia



Northwestern Digital Image Lib.

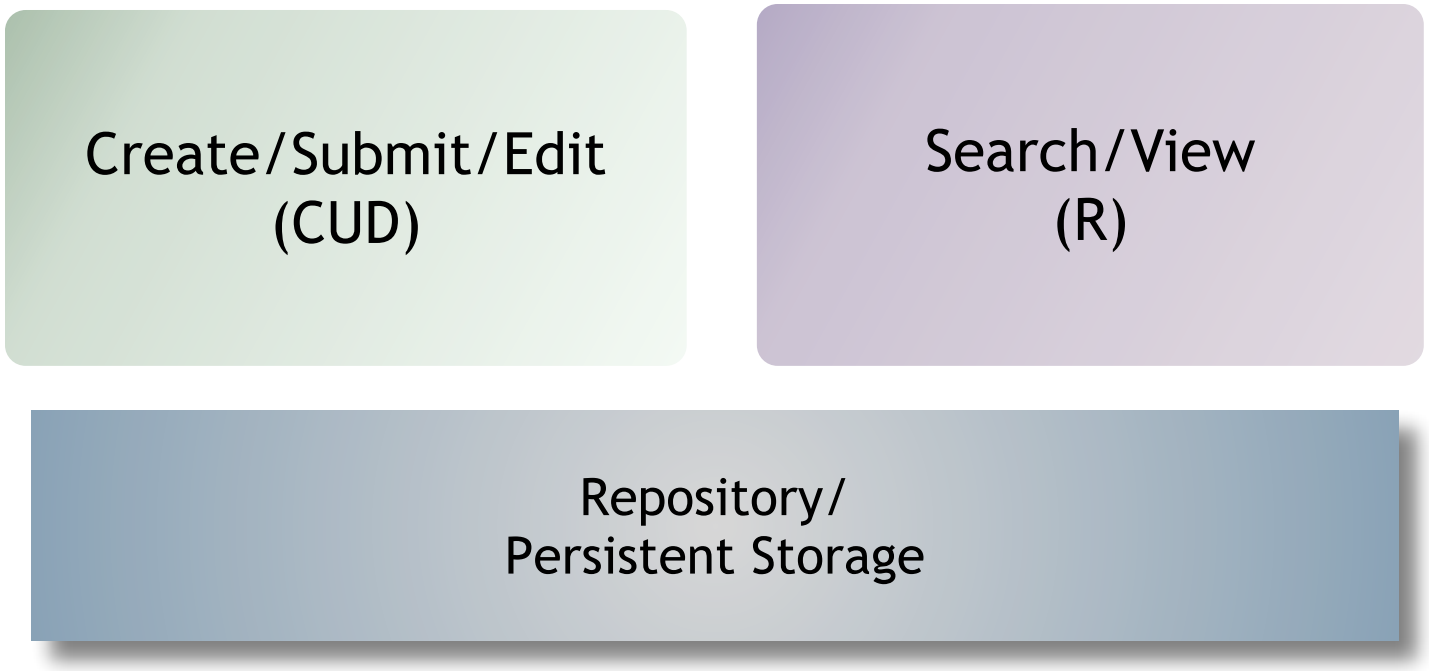
Hydra Partners *and Known Users*



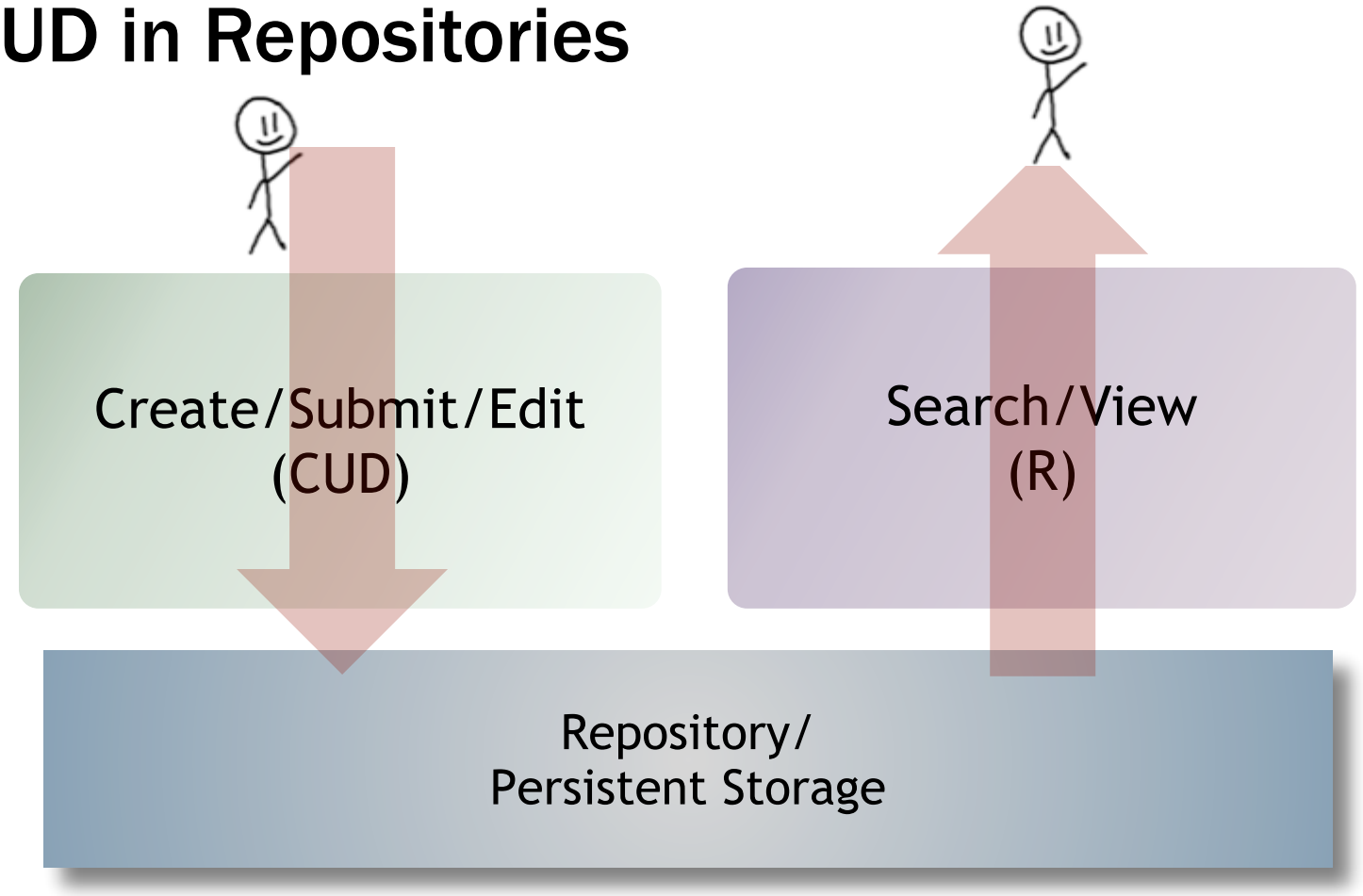
OR = Open Repositories Conference

Hydra Technical Framework

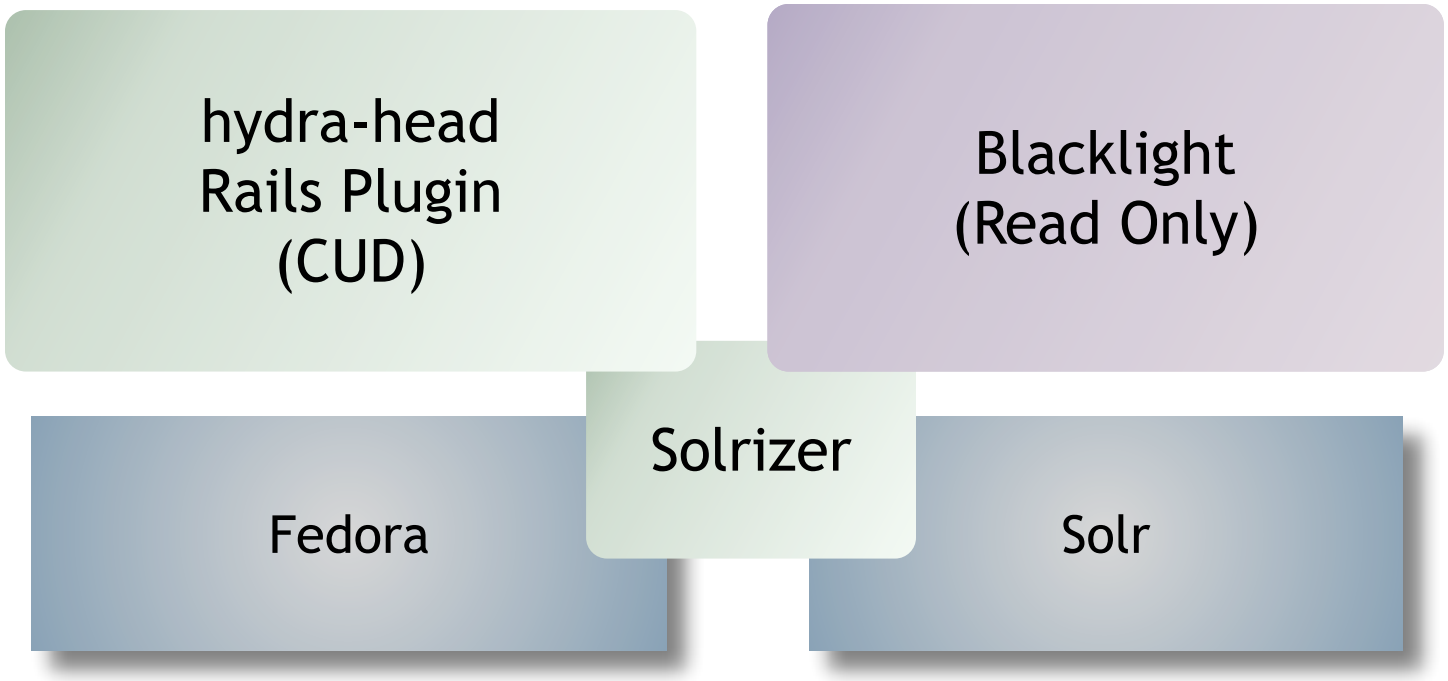
CRUD in Repositories



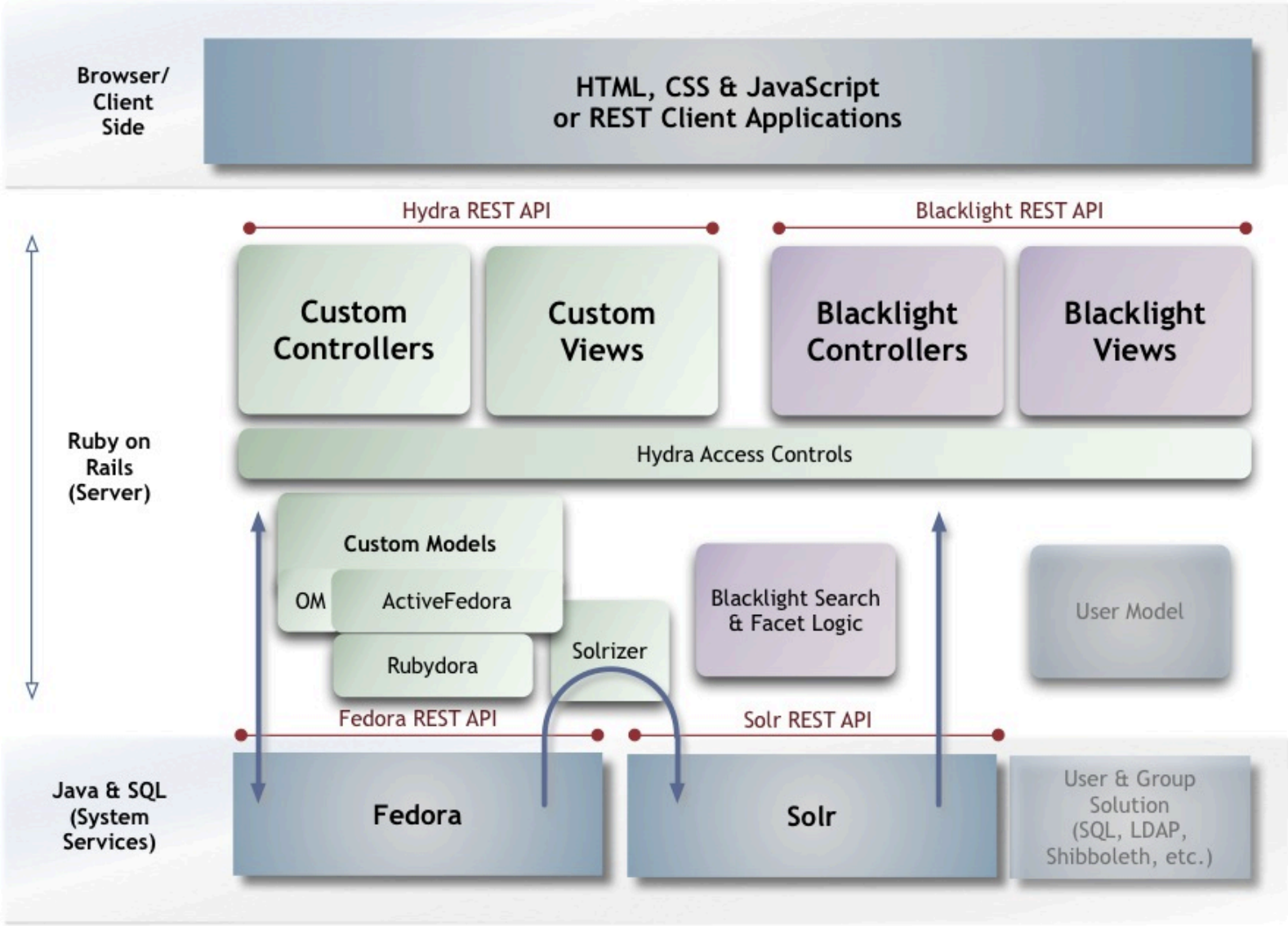
CRUD in Repositories



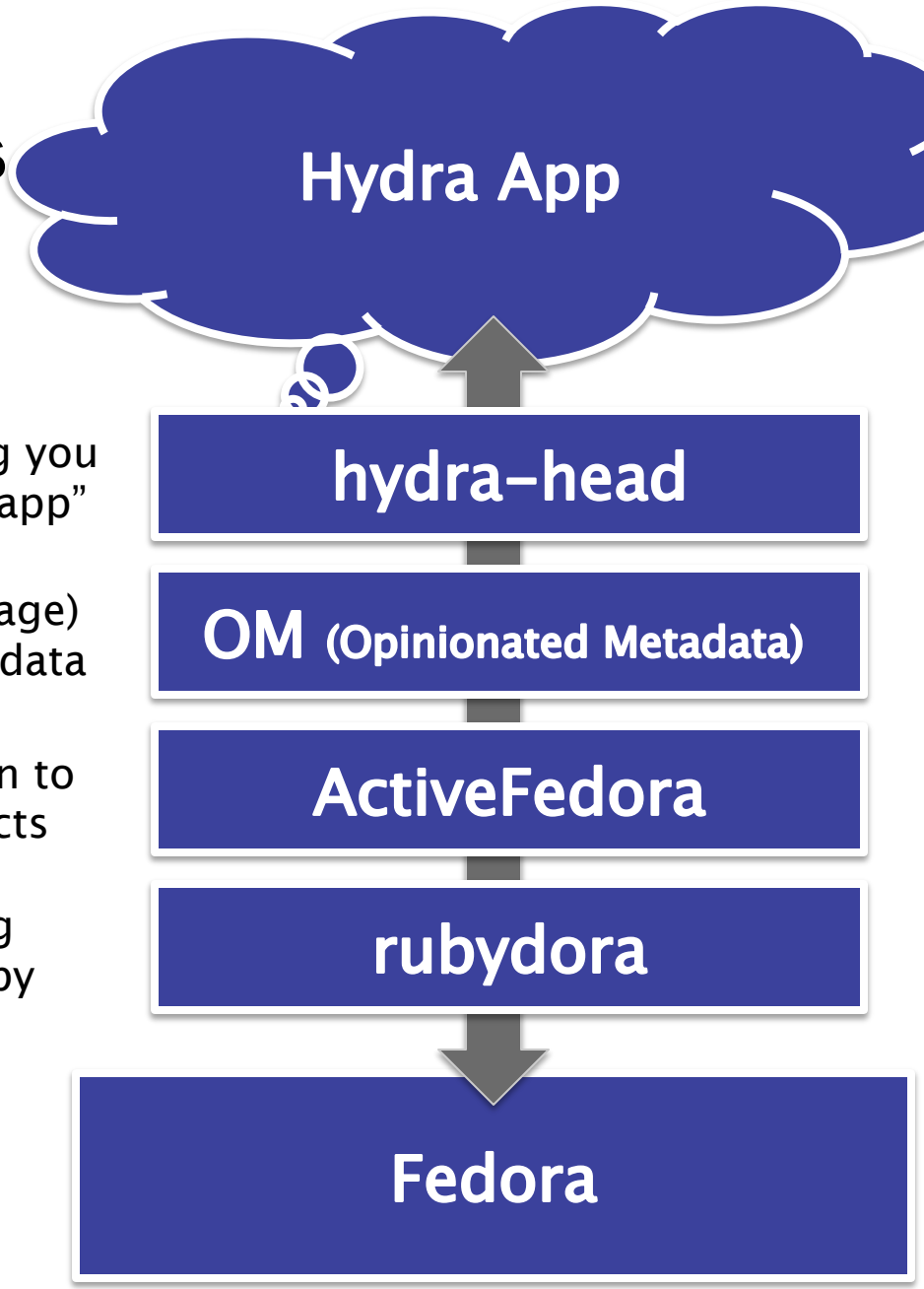
Major Hydra Components



The Hydra Stack



Hydra's Core Gems



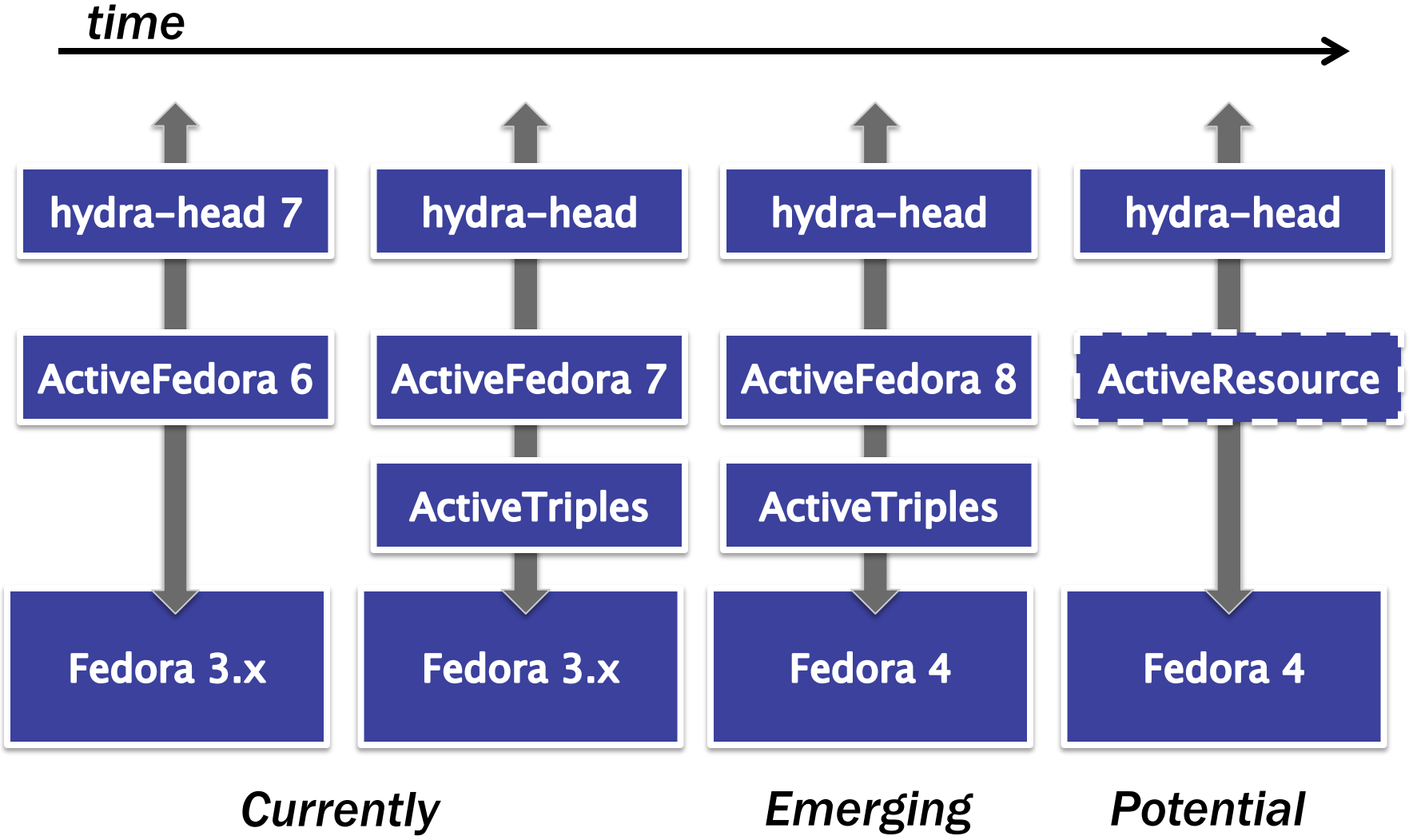
Rails engine with "everything you need to build a user-facing app"

DSL (domain specific language) for easy mgmt of XML metadata

Applies ActiveModel pattern to working with Fedora objects

Client API for consuming Fedora's REST API via Ruby

Hydra's Core Gems

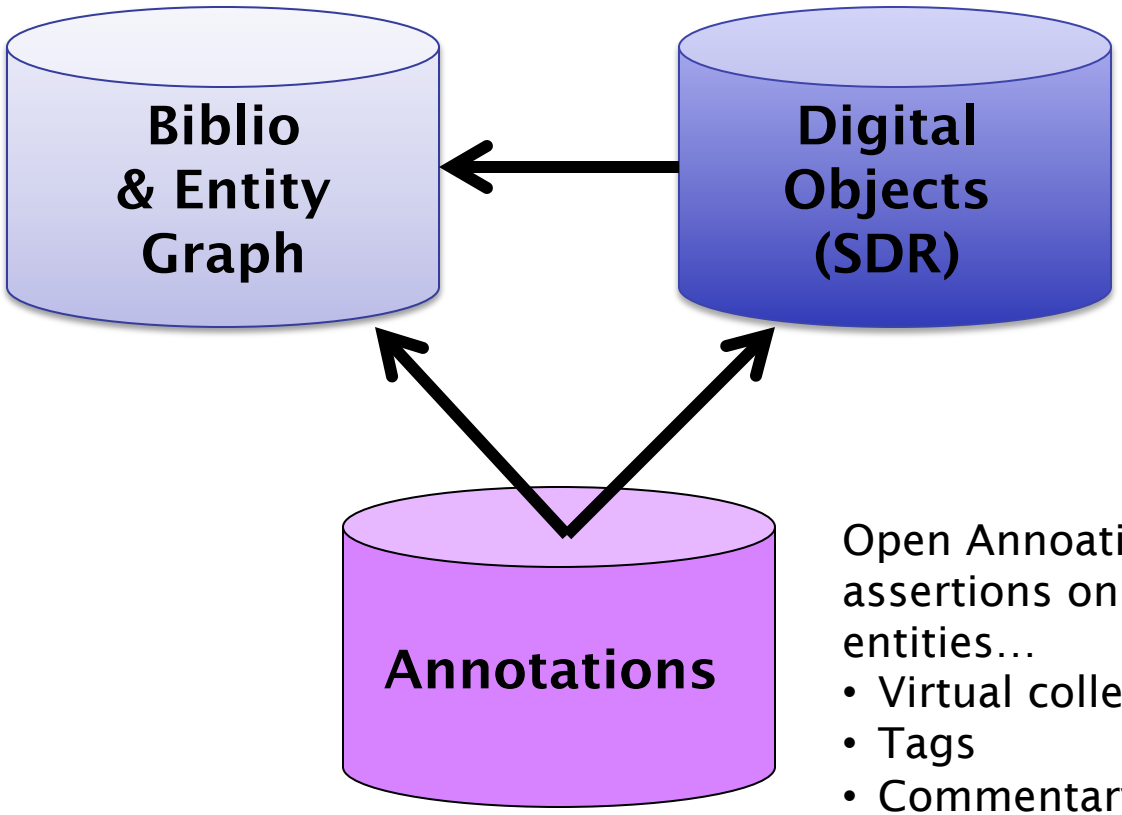


Three Active Hydra-Fedora 4 Pilots

- Penn State University / ScholarSphere
- UCSD DAMS
- Stanford / Annotation Management

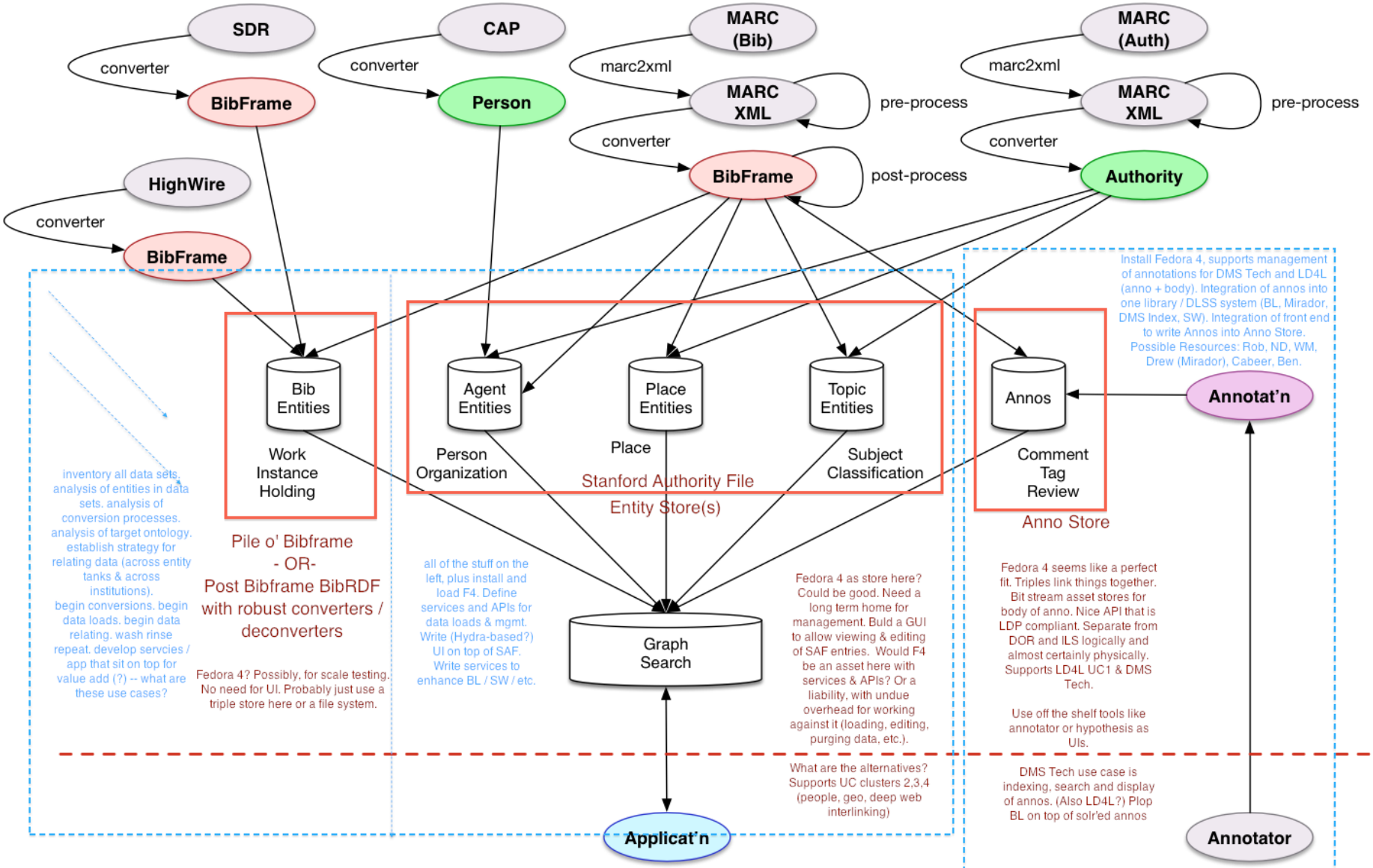
Project Cerberus at Stanford

- Bib records (Bibframe)
- Local authority file (person, org, place, subject, titles) ,



- Books
- Images
- Manuscipts
- Media
- Software
- Etc.

Open Annoation-based assertions on other entities…

- Virtual collections
- Tags
- Commentary
- Translation
- Etc

Project Cerberus at Stanford