Integration Challenges & Rewards:
Heterogeneous Solutions with Fedora4 at the Core

UNIVERSITY OF VIRGINIA LIBRARY

Ellen C. Ramsey, University of Virginia, ellen@virginia.edu
Robin Ruggaber, University of Virginia, robin.ruggaber@virginia.edu

Discussion Points

- Drivers for How We Solve Problems
- Three Challenging Use Cases
- Integration Challenges & Costs in Heterogeneous Environment
- Architecture with Fedora at the Core
- Q & A

Meet the Need, Beat the Clock or Lose the Business

Image Credits: drjill-live.com

Don’t Drop the Ball

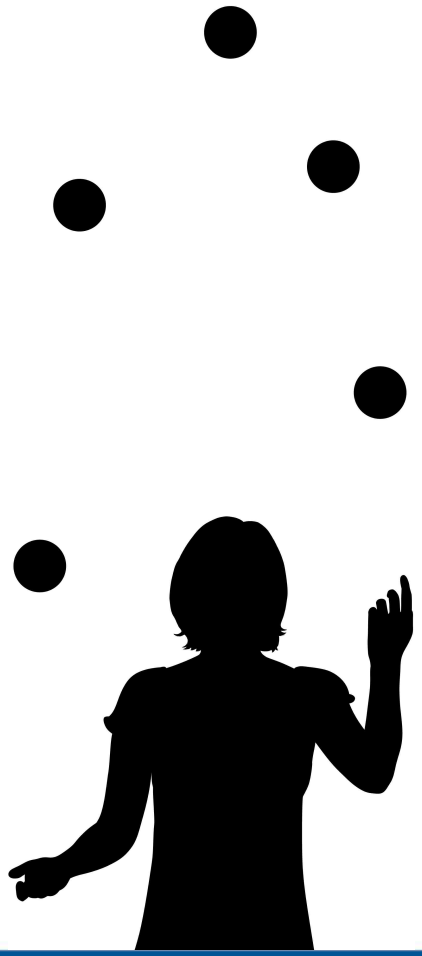


Image Credits: Mo O’Hara

Reduce Risk

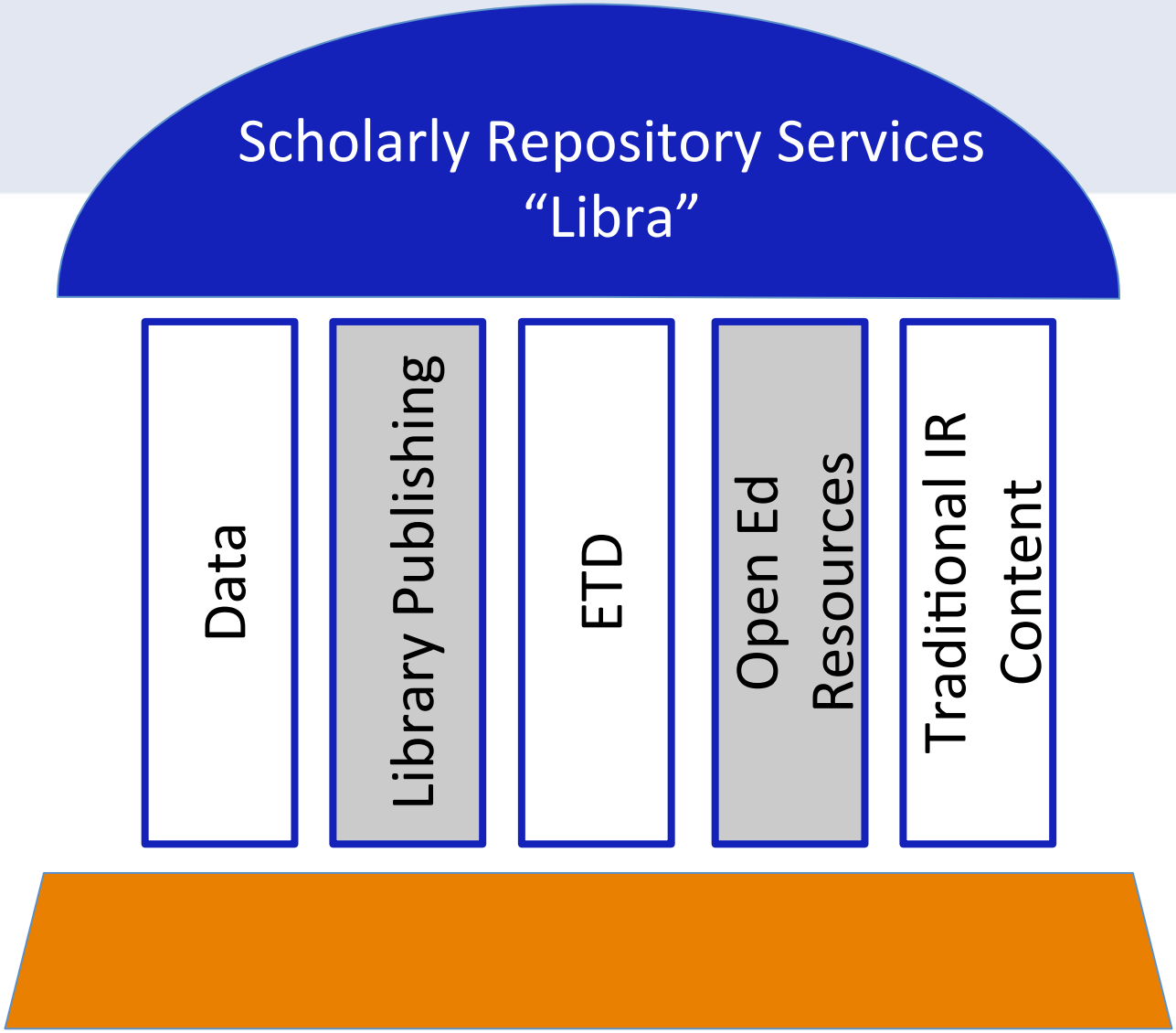
Software	Hardware	Content
Community Driven Development	Secured Access	Distinct Copies Stored to Avoid Common Threats
Open Source	Highly Available, Redundant Storage	Secure Access
Managed, Tested Code	Clustered Compute	Fixity
For the Academy, By The Academy	Minimize Single Points of Failure	Version Control
		Persistent Identifiers

A Few Choices We Have Made

- Global Search Across All Content
- Community Driven Opensource Software e.g.
  - Fedora
  - Solr
  - Blacklight
  - Hydra
    - Avalon
    - GeoBlacklight
    - Sufia
  - ArchiveSpace
- Consortia Based Digital Preservation
  - APTrust
  - DPN
- Interoperable Components
  - IIIF image server & viewers
  - All of the Above

Challenging Use Cases

1. Scholarly Services - Research Data
2. At Risk Web Content
3. Cultural Heritage - The Renovation of Thomas Jefferson's Rotunda


Scholarly Repository Services
"Libra"
Data
Library Publishing
ETD
Open Ed Resources
Traditional IR Content

Envisioning a New Libra

Choices

Why is a Hydra institution choosing things that don't glue into the Hydra infrastructure?

1. It was about time and features needed.
2. Had to have a solution for grant required public data sharing.
3. Environmental scan results pointed away from Hydra (for now).

Drivers

- Nationwide resurgence of OA mandates
- OSTP/Federal government requirements
- Contributing to open source development communities

Stakeholders

Concerns

1. VPR/IRB - research & researcher compliance, OSF Integration
2. Office of Sponsored Programs - Funder compliance
3. University IT - authentication, authorization and data security
4. Local Researchers - easy to use, easy to find
5. Other Researchers - easy to re-use
6. Library - preservation, discoverability, dissemination, interoperability, stability & sustainability of solutions selected

Search for a Research Data Solution

Evaluation Criteria

Four broad categories of requirements were applied to data repository candidates:

1. Faculty needs/Funder compliance
2. Statistics and reporting
3. Security (Authorization/Authentication)
4. Architectural Interoperability
5. Metadata

Contenders

- Dataverse
- Sufia (data aspects)
- DSpace

Selection is...Dataverse

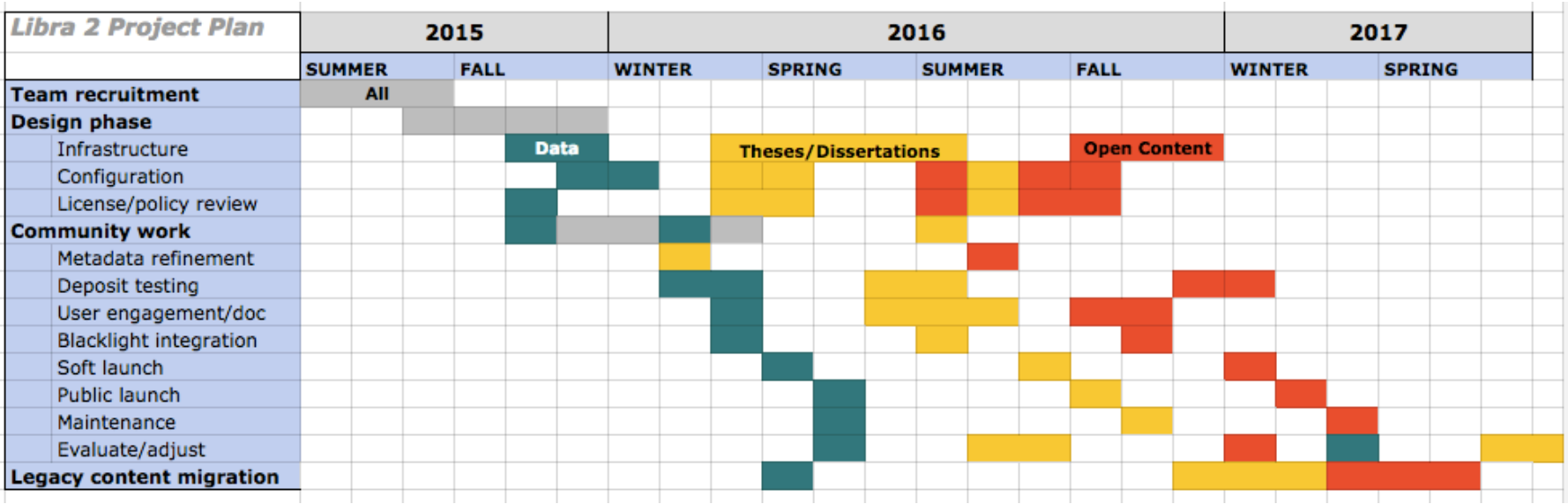
Benefits

1. Speed of Delivery
2. Fit user needs (results of dataverse config user tests)
3. Produces SOLR indexing records for ease of global search
4. Interoperability with Open Science Framework
5. Known to users

Compromises

- Open source but limited adoption/development
- Doesn’t interoperate with Fedora out of the box, APIs evolving
- Not purpose built (turn off collaborative functions)

Libra 2 Development



Challenging Use Cases

1. Scholarly Services - Research Data
2. At Risk Web Content
3. Cultural Heritage - The Renovation of Thomas Jefferson's Rotunda

At Risk Web Content
Web Archiving with Archive It

Gains:

- Technology to crawl at risk websites & develop WARCS
- Processing of preservation & descriptive metadata
- API allows retrieval of the WARCS and preservation metadata
- Dissemination of sites through Internet Archive

Challenges:

- No descriptive metadata API
- Minimal preservation metadata

Approach:

- Testing API for retrieval of WARCs & associated metadata
- Consulting to improve preservation metadata
- Informing needs for API for descriptive metadata & ready to test

Challenging Use Cases

1. Scholarly Services - Research Data
2. At Risk Web Content
3. Cultural Heritage - The Renovation of Thomas Jefferson's Rotunda

Capturing the Renovation of Thomas Jefferson's Rotunda Partnering with Cyark

Gains:

- Technology to process multi-point scans to produce images and guidance
- Dissemination of images through Cyarks showcase of world heritage sites

Challenges:

- Exchange of content is physical
- No APIs for metadata
- No path to APTrust

Approach:

- We plan to store datasets in Dataverse to enable discovery and access
- Continue to work with Cyark

Fedora4 At the Core

1. As a Linked Data Platform – It Supports the Present and the Future We Want
2. API Approach – Flexible, Interoperable
3. Sustainable Approach – Community Driven Development
4. Foundation for Hydra! One Body, Many Heads…Plus

Questions?