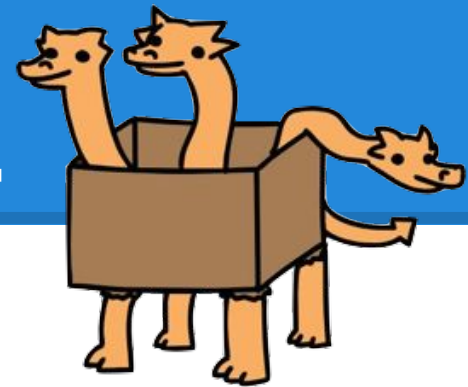# Hydra in a Box
Building and Bundling a National Digital Platform

Tom Cramer (Stanford University)
Debra Hanken Kurtz (DuraSpace)
Mark Matienzo (Digital Public Library of America)
CNI Membership Meeting — December 15, 2015

Background

A Time for Change

- Conversations between Stanford University, DPLA, and DuraSpace informed project design

- Current digital collections platforms originate in an earlier phase of the web, which explain current limitations

Infrastructure needs in the network of DPLA and its Hubs

- Legacy systems unable to leverage modern affordances of the web

- Lack of scalable and sustainable aggregation workflows

- Lack of support for linked data and metadata enrichment

- Perceived lack of "obvious choices" for replacement systems for digital collections
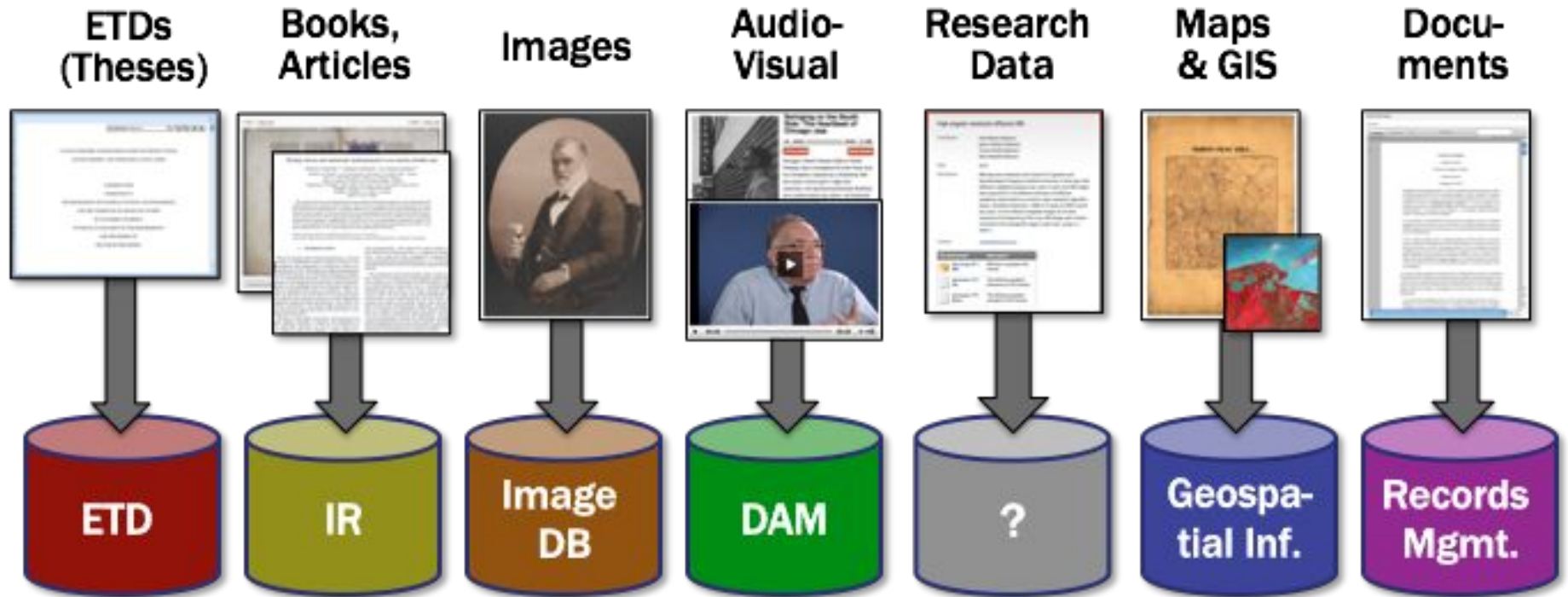
The Vision

- A product and service that is easy to use, easy to integrate, and that

- Reduce barriers (including cost) to DPLA contribution

- Allow digital collections to be not just *on* the web, but *of* the web

- Expand and diversify both the DPLA and Hydra communities

Current State of Hydra

What is Hydra?

- A robust repository fronted by tailored applications and workflows ("heads")
  - *One body, many heads*
- Collaboratively build "solution bundles"
- A community of partners extending & enhancing the core

  - *If you want to go fast, go alone. If you want to go far, go together.*

A Common Approach

A Repository Approach

ETDs (Theses)

Books, Articles

Images

Audio-Visual

Research Data

Maps & GIS

Docu-ments

Digital Repository

Scalable, Robust, Shared Management and Preservation Services
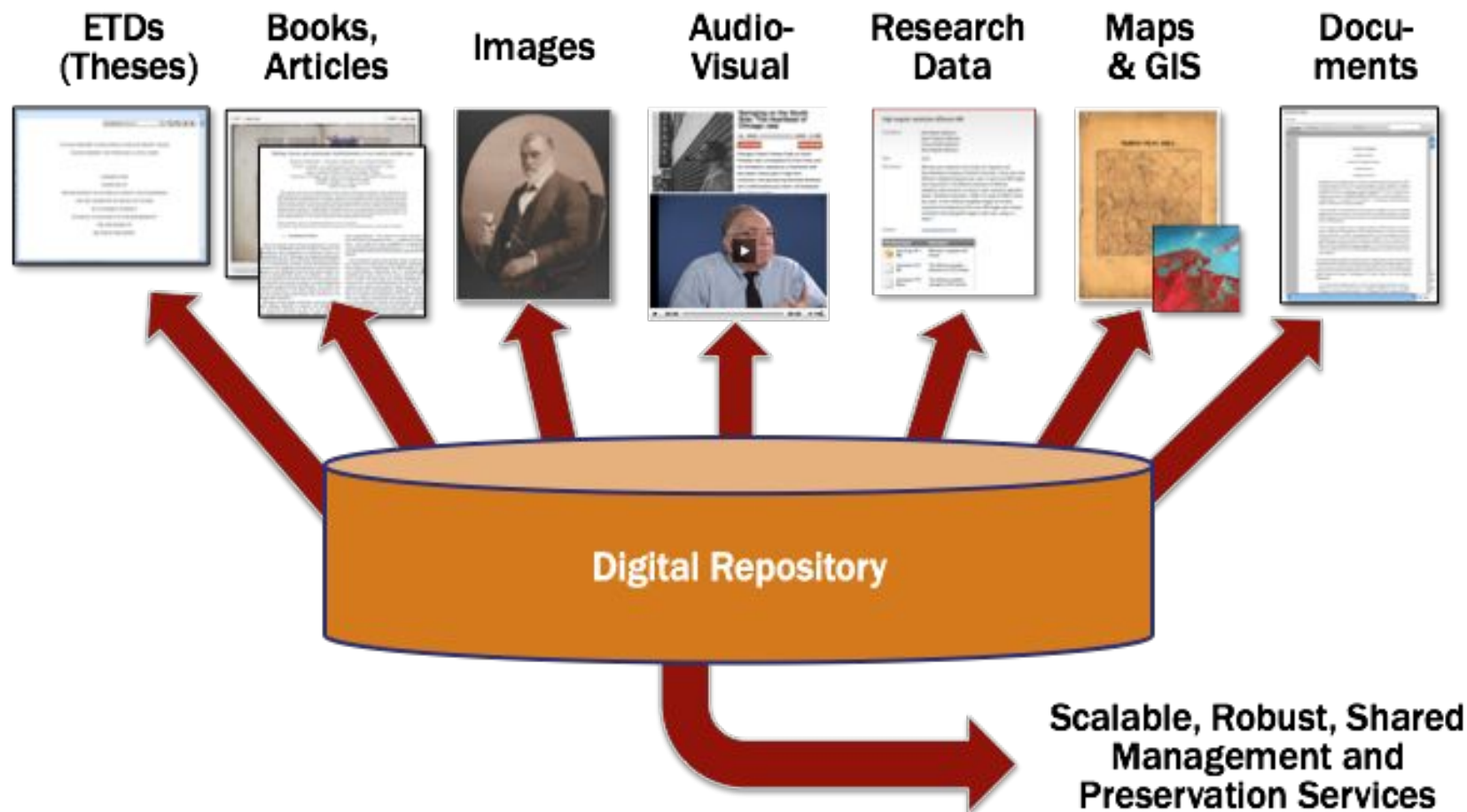
One Body, Many Heads

Hydra Partners & Users

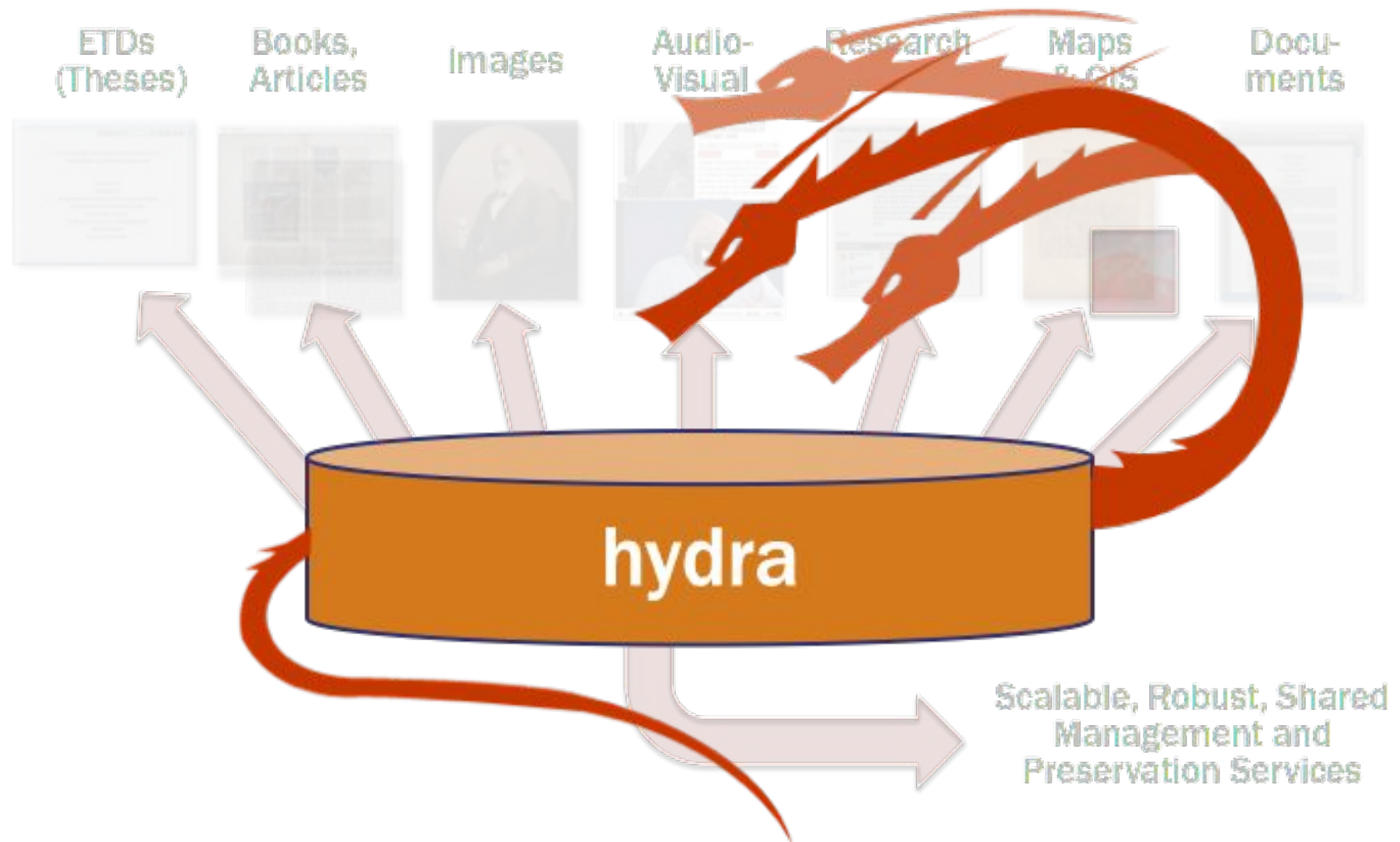- Stanford University
- DuraSpace
- Columbia University
- London School of Economics
- Data Curation Experts
- Duke University
- University of Cincinnati
- University of Oregon (Oregon Digital)
- Tufts University
- Digital Public Library of America

- University of Hull
- University of Notre Dame
- Penn State University
- Rock and Roll Hall of Fame
- WGBH
- Yale University
- Princeton University Library
- Oregon State University (Oregon Digital)
- Duoc UC
- University of Michigan

- University of Virginia
- Northwestern University
- Indiana University
- The Royal Library of Denmark
- Boston Public Library
- Virginia Tech
- Cornell University
- Case Western Reserve University
- University of Alberta
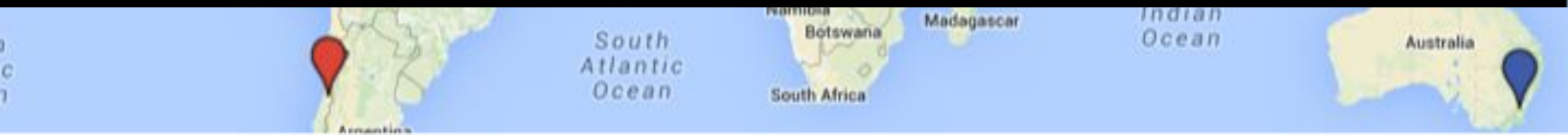
Hydra Partners and *Known* Users
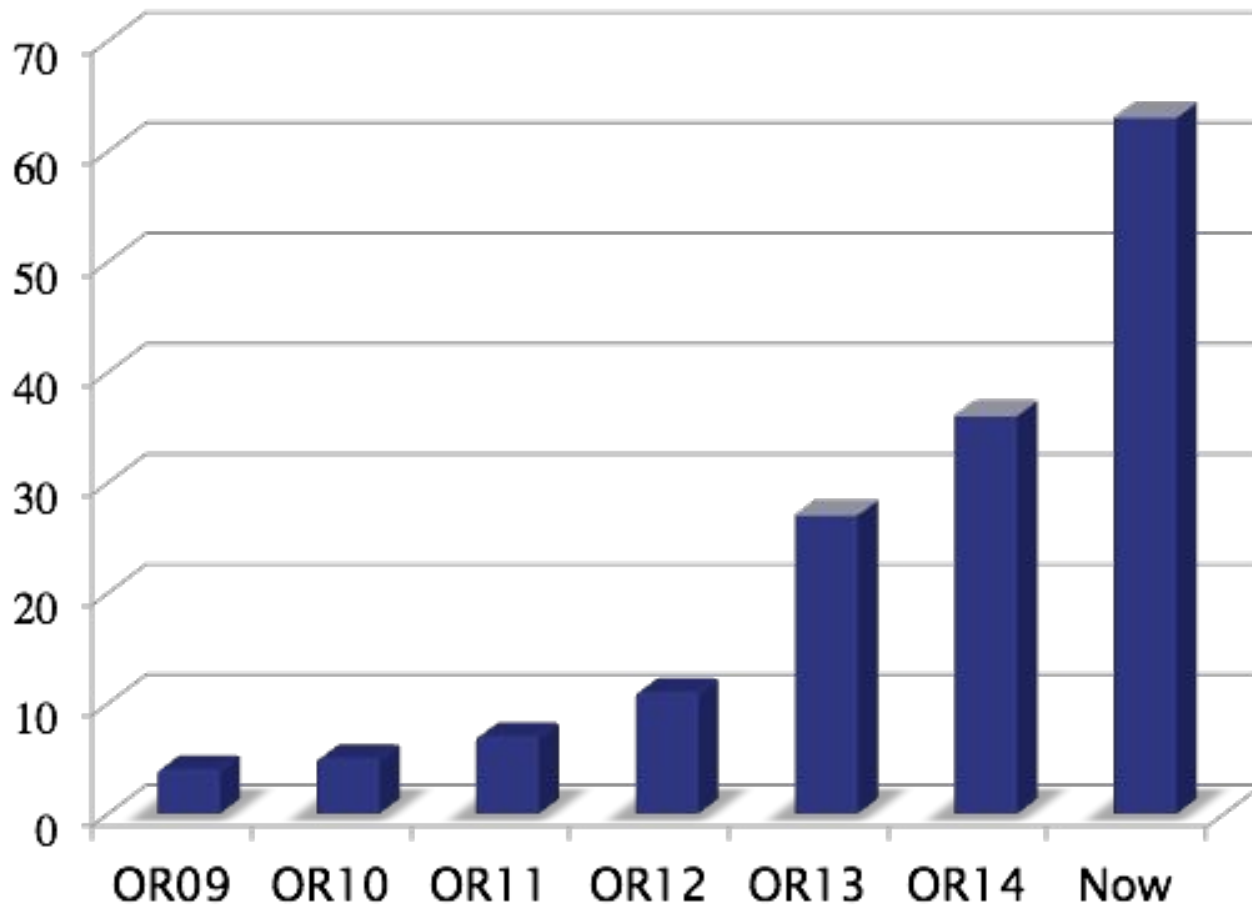
Hydra Partners, Known Users and Licensed Contributors
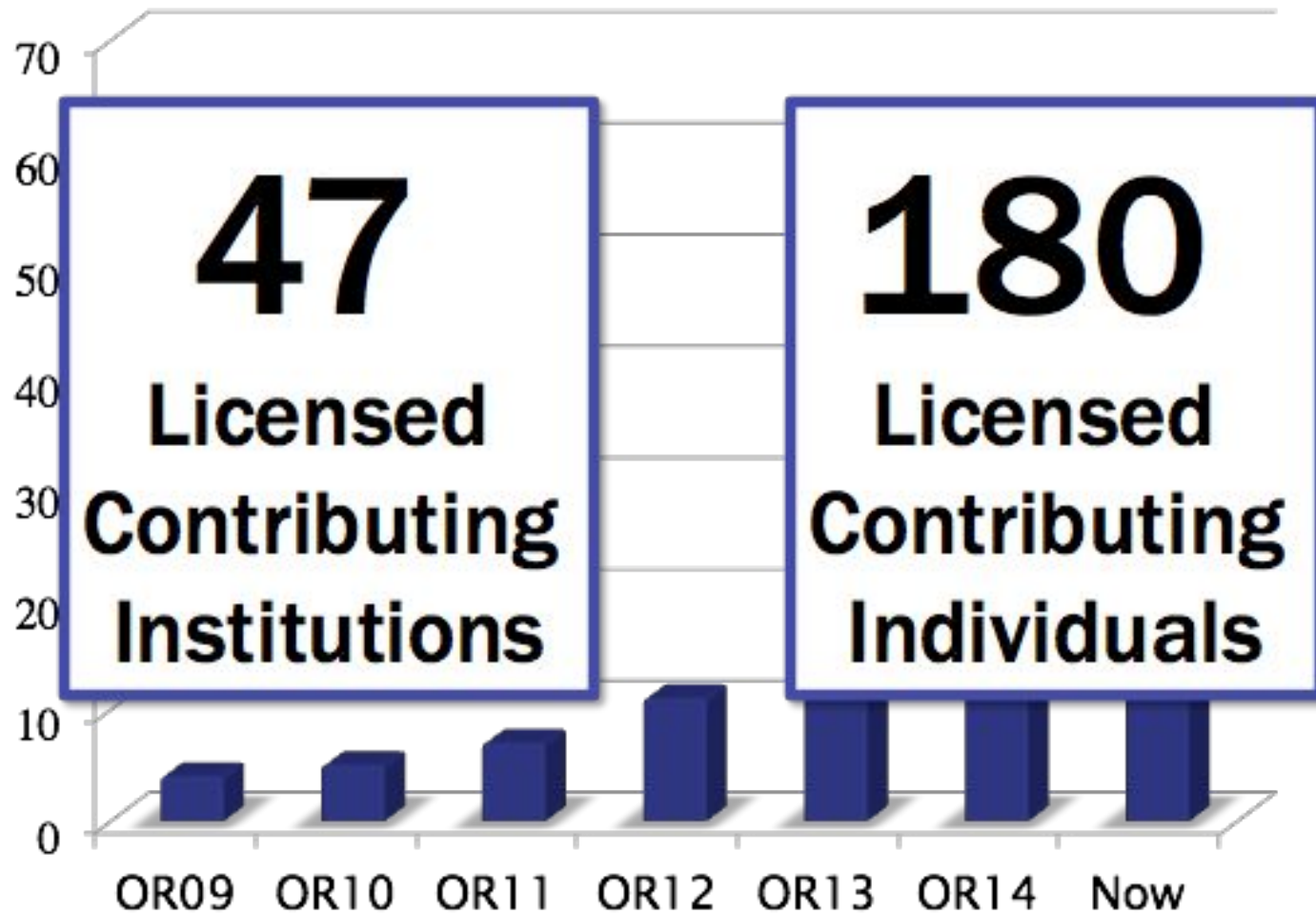
Going Far Together

3rd Worldwide Hydra Connect @ Minneapolis, OH, September 2015
200 participants from 60 institutions

Interest & Working Groups

- Fedora 3 to 4 Migration Interest Group
- Hydra Archivists Interest Group (HAWG)
- Hydra DevOps Interest Group
- Hydra Digital Preservation Interest Group
- Hydra Geospatial Interest Group
- Hydra GIS Data Modeling Working Group
- Hydra Marketing Interest Group
- Hydra Metadata Interest Group (formerly Working Group)
- Hydra Metrics Interest Group
- Hydra Service Management Interest Group
- Hydra Time-based Media Interest Group
- Hydra Triple Store Interest Group
- Hydra User Experience Interest Group
- Hydra Web Presence Working Group
- Sufia User Interface Working Group
- Sufia Documentation for Managers Working Group

https://wiki.duraspace.org/display/hydra/The+Hydra+Project

Used By....

- Large Universities
- Small Universities
- Colleges
- Public Broadcasting
- Government Ministry
- National Libraries
- National Lab
- Small Research Labs
- National Digital Repository
- Statewide Digital Libraries
- Rock & Roll Hall of Fame
- Chemical Heritage Foundation
- Museum of Performing Arts
- A Shakespeare Festival

Used for...

- Self-deposit system
- Digital Collections System
  - Sheet Music
  - Architectural resources
- Electronic Theses & Dissertations
- Digital Image System
- Media Management
- Media Preservation System
- Research Data Management
- Digitization Workflow System
- Digital Preservation System
- Digital Archives System
- Library Monograph Acquisitions System
- And more!

The Hydra Way

- Not a directed project
- Investment in a framework, not an application
  - Contributions back to core code base
- Investment in a community, not a vendor
  - Contributions back to community: training, documentation, modeling, sharing best practices, outreach
- Travel & Face time

Major Technical Advances

How Does Hydra Scale Next?

We need a solution

  ...that is fully featured

    ...easy to install & maintain

      ...concentrates technical dev.

        ...gives an on-ramp for new sites

How Does Hydra Scale Next?

We need a solution

...that is fully featured

...easy to install & maintain

...concentrates technical dev.

...gives an on-ramp for new sites

We need to build and bundle.

How Does Hydra Scale Next?

We need a solution

   ...that is fully featured

      ...easy to install & maintain

         ...concentrates technical dev.

            ...gives an on-ramp for new sites

We need to build and bundle.

We need a box.
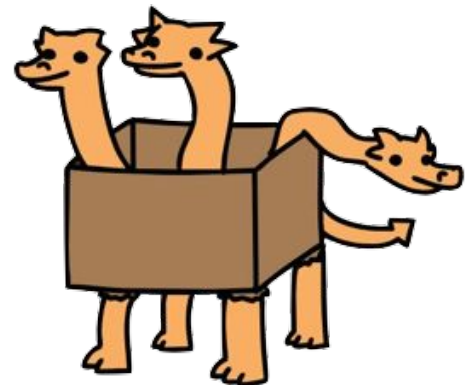
Project Overview

The Pitch

"The right set of tools for managing, publishing, and sharing digital collections, marshaled by the right collaborators, has the potential to make the dream of a national digital platform for cultural heritage institutions a reality."

Scope

"...a directed project, with a unified and integrated project team spanning three different organizations and multiple locations."

Deliverables

- Create a turnkey, Hydra-based application
- Improve and generalize DPLA's metadata ingestion system into an "aggregator system in a box"
- Connect these key infrastructural pieces with DPLA hubs, current Hydra partners, and prospective Hydra adopters
- Offer a cloud-based version of the application for use across multiple domains

Hosting

HydraDirect
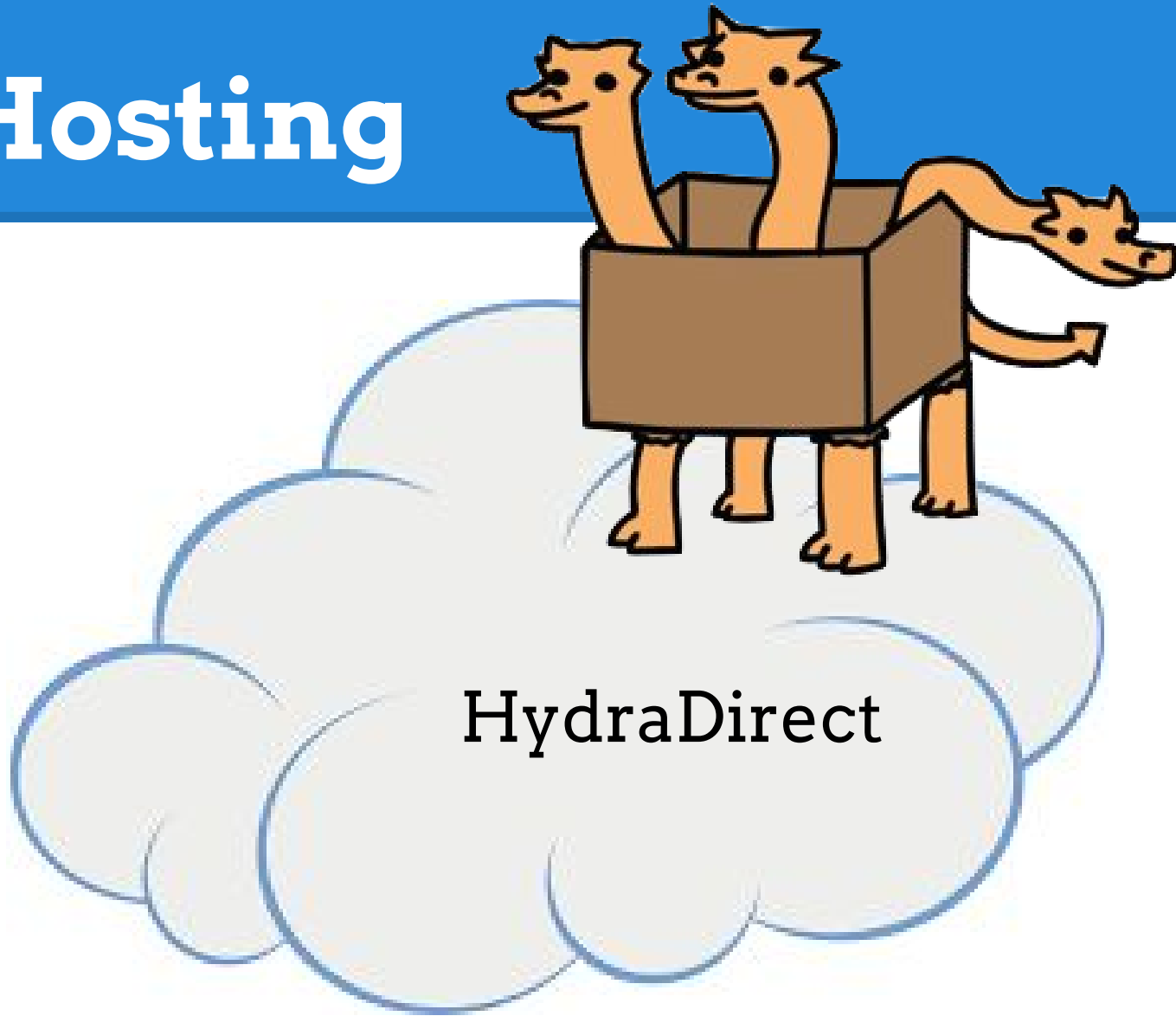
What's the timeline for the project?

Overall timeline

- May 2015-November 2017 (30 months)
- Design process: May 2015-February 2016
- Development: March 2016-Nov. 2017
- Service development and community engagement: throughout project

Design Process

- Discovery Phase (Summer - Fall 2015)
  - Literature review and product/service analysis
  - Surveys, interviews, and focus groups
  - Community outreach
- Technical exploration (Fall/Winter)
- Information Architecture (Winter)
  - User requirements and personas
  - Requirements - functional & technical
  - Models
  - Wireframes
- Visual Design (Early spring)

Hydra-in-a-Box Design Process: Tasks & Timeline

Updated 2015-09-18

Discovery

Phase 1 → **Phase 2**

Stakeholder goals
Community & landscape surveys

Interviews & focus groups

Information Architecture

Phase 1 → **Phase 2**

User personas
Requirements prioritization

Conceptual sitemaps
Wireframes

Visual Design

Phase 1

Visual design mockups
Visual design style guide

Infrastructure & Technical Exploration

Phase 1 → **Phase 2** → **Phase 3**

Community engagement
Collaborative development on core gems

Technical prototyping

Gap analysis
Community feedback

Development

Phase 1 →

Formal development & user testing

Discovery: Phase 1	Discovery: Phase 2	Info Arch: Phase 1	Info Arch: Phase 2
	Tech Exploration: Phase 1	Tech Exploration: Phase 2	Tech Exploration: Phase 3

August	September	October	November	December	January	February	March

2015   2016

Progress

Key progress so far

- Landscape analysis: survey, literature review, product and service analyses

- Deeper feedback and gathering requirements: interviews, focus groups

- Identification of requirements for content types

- Creation and charging of subteams for focused work areas

Survey

256 complete responses

311 repositories

Mostly small, US academic libraries

Public college/university library
34%

Independent research library/archives
14%

Museum
10%

Over half of respondents represent either a public or private college or university library.

Regional consortium
8%

Historical society
4%

Private college/university library
20%

Public library
6%

Government
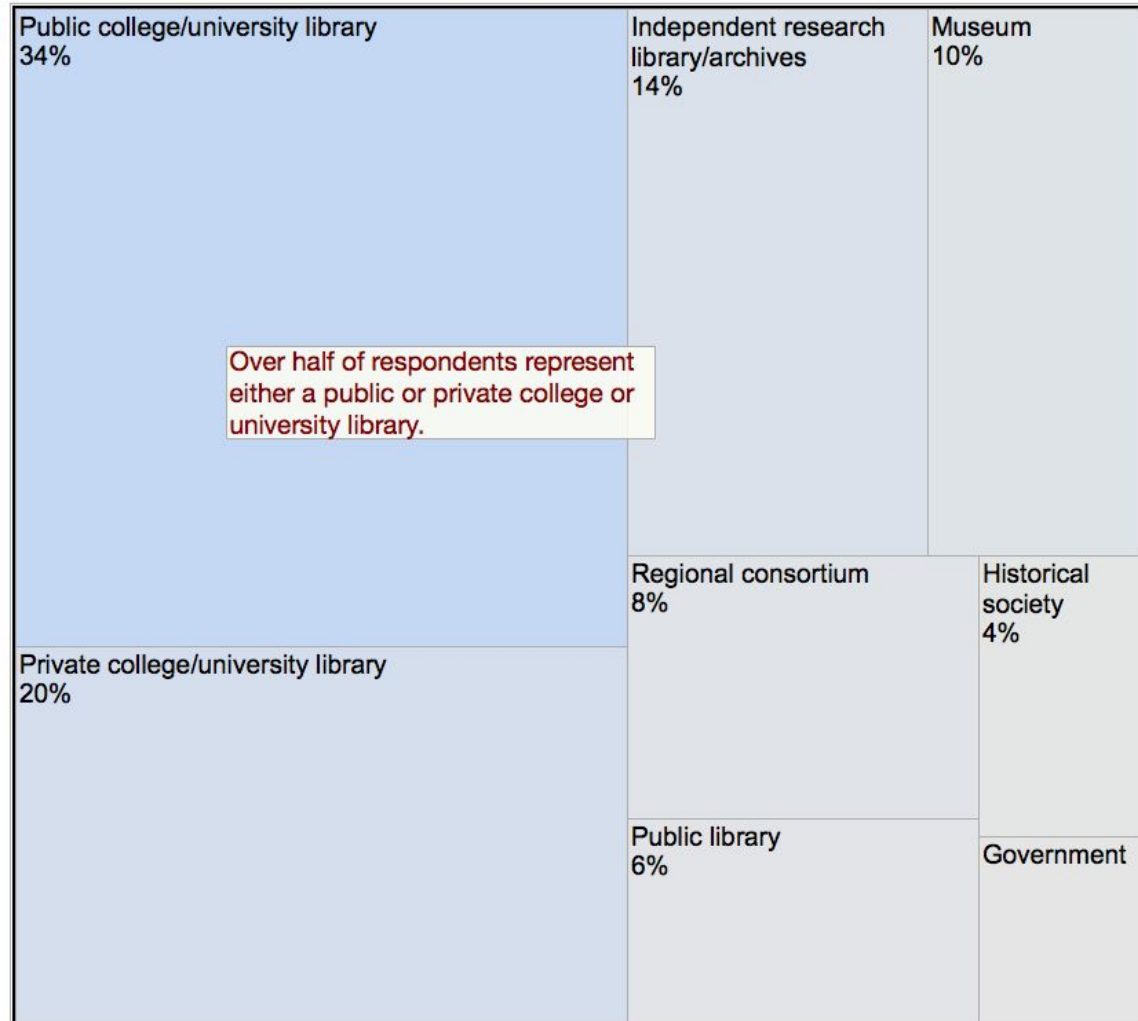
Survey Insights

- Expectations
- Satisfaction levels
  - Users of hosted services tend to be more satisfied than users of local deployments
- Strengths and weaknesses of existing repository options
- 53% plan to migrate to another system
  - Most to a Fedora-based solution
  - Rest are "not sure" what's next

Interviews and focus groups

- Completed 21 individual or small-group interviews and 4 focus groups

- Interviews held either in-person or through videoconference; focus groups held in-person

- Undergoing coding and analysis process to further identify potential requirements

Content type analysis

	IMAGES									TEXT	
	digital photo	digital photo album	scanned poster	scanned postcard	scanned looseleaf manuscript	scanned folder of archival docs	scanned photo album (pages)	scanned scrapbook (pages and clippings)	scanned newspaper	scanned monograph with cleaned OCR	scann mono with
OBJECT STRUCTURE											
single file	yes	no	yes	no	no	no	no	no	no	no	no
multi-file, sequenced, flat	no	yes	no	yes	yes	yes	yes	yes	yes	yes	yes
multi-file, unsequenced, flat	no	no	no	no	no	no	no	no	no	no	no
multi-file, hierarchical	no	no	no	no	no	no	no	yes	yes	no	no
DESCRIPTIVE METADATA											
basic ("core")	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
geo	yes	yes	no	yes	no	no	no	no	no	no	geo
serial (volume, issue, etc.)	no	no	no	no	no	no	no	no	yes	no	seria
archival order/hierarchy	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no
DELIVER DERIVATIVES	maybe	no	yes	yes	yes	yes	yes	yes	yes	yes	yes
APPLICATION BEHAVIOR (ITEM LEVEL)											
zoom	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
front/back	no	no	no	yes	yes	yes	no	yes	no	no	fron
page turn	no	no	no	no	no	no	yes	yes	yes	yes	yes
image set (gallery)	no	yes	no	no	no	yes	yes	yes	yes	no	ino
media preview (storyboard, poster)	no	no	no	no	no	no	no	no	no	no	no
calendar based search/browse	no	no	no	no	no	no	no	no	yes	no	no
text transcription	no	no	no	no	maybe	maybe	no	maybe	yes	no	no
full text search	no	no	no	no	maybe	maybe	no	maybe	yes	yes	no
index map	no	no	no	no	no	no	no	no	no	no	index
feature extraction	no	no	no	no	no	no	no	no	no	no	no

Subteam development

- Charging dedicated sub-teams for specific workstreams
  - Design, requirements and specifications
  - Data modeling
  - Service development (separate tech and business teams)
  - Marketing and communications
- Clear identification of responsibilities because of distributed nature of project team

Additional information

Want to know more?

How to stay informed

Visit our website and blog:
http://hydrainabox.org/

Public information list
hybox-info@googlegroups.com

Contact us
hybox-contact@googlegroups.com

How to participate

- Contribute ideas for functionality: **http://bit.ly/hydrainabox-ideas**

- Help us test and evaluate the software as we release it

- Participate in software development

Thank you!

Tom Cramer

tcramer@stanford.edu

@tcramer

Debra Hanken Kurtz

dkurtz@duraspace.org

@debra

Mark Matienzo

mark@dp.la

@anarchivist