

# Scale

## Long-term Digital Asset Management Initiative at the University of Wisconsin-Madison

Lee Konrad

Associate University Librarian for Technology, Data, and Digital Library Services

[Lee.Konrad@wisc.edu](mailto:Lee.Konrad@wisc.edu)

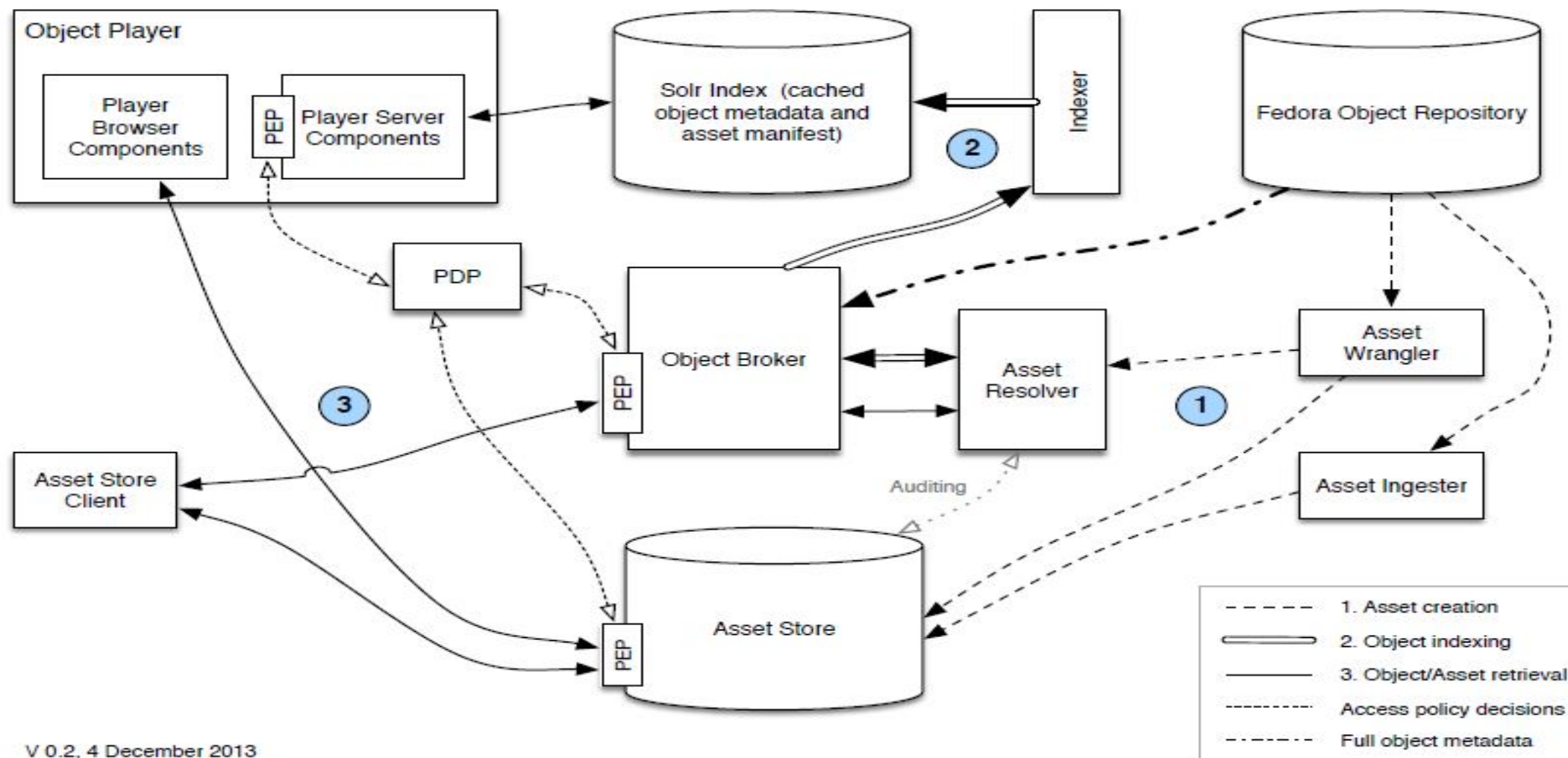


# University of Wisconsin

- UW System
  - Two doctoral institutions A dozen comprehensives
  - Multiple library DL and IR environments
- University of Wisconsin-Madison
  - Libraries, Archives, Museums, Research Centers, Administrative, etc.
- UW Digital Collections
  - Content contributors
    - 20 years of digital collection building – 100+ disparate collections – 3M objects
    - UW Libraries, Archives, State agencies, public libraries (DPLA), campus partners, etc.
  - Digital Ecosystem
    - Digital library, IR, licensed content, etc.
    - 100 TB of material warranting preservation
    - 10M uses per year
    - Redundant and geographically distributed storage
    - Fedora
    - Dspace
    - Coordinated Discovery

# Digital Collections Infrastructure

Public components status - June 2014



V 0.2, 4 December 2013

# Selected Digital Collection Models

Collection Characteristics Igko

File Edit View Insert Format Data Tools Add-ons Help Last edit was made on March 15 by PETER GORMAN Comments

€ % .0 .00 123 Verdana 10 B I A

fx Last updated:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Last updated:	2017-02-09										
2												
3	citation only											
4	restricted access											
5												
6	Collection ID	citation	ImBib (simple)	ImBib (complex)	EFacs	TEIU	EAD	HTML	other	Access	Activity	notes
23	Primate		X							open		
24	SouthAsiaIndex	X								open		index nonroman characters
25	UWDCCitation	X								open		citations for Efac items when collection has no MmBib component
26	WAVE	X								open		
27	WiscFolkSong	X		X						open		
28	WiscSheetMus	X		X						open		
29												
30	AfricanStudies		X		X					open		
31	AldoLeopold		X		X					open		
32	DLDecArts		X		X					open		HTML in metadata - Tier 4?
33	Dominy			X	X					open		
34	EastAsian		X		X					open		
35	EastEurope		X		X					open		
36	EcoNatRes			X	X					open		
37	FRUS				X					open		
38	GenderStudies		X		X					open		
39	German		X		X					open		
40	HistSciTech				X					open		
41	HumanEcol			X	X					open		
42	JoyceColl				X					open	static	
43	Preservation				X					open		
44	RealEstate				X					open		
45	Science		X		X					open		
46	SEAIT		X		X					open		
47	SilverBuckle				X					open		
48	SLISTeaching		X		X					open		
49	TamsWitmark			X	X					open		
50	UWBoR				X					open	static	
51	UWSpecColl			X	X					open		
52	Zooloqv		X		X					open		

# UW-Madison – Preservation Pilot

- Clear need within the Libraries and Archives divisions
- Clear risk to the institutional investments in digital collection building
- Presumed need with respect to other campus stakeholders / entities
- Excellent central backup IT services – no actual preservation class services offered on campus
- Ad hoc working group began considering the needs and risks – formalized as LDAM
- Hitachi Content Platform – Joint project of campus IT and Libraries

# UW-Madison and DPN

- DPN – Charter Members
- DPN membership jointly sponsored by
  - Vice Provost for Libraries
  - Vice Provost for Information Technology and CIO
- Campus partners
  - Libraries and Archives
  - CIO's office
  - Division of Information Technology
  - Legal Services and Business Services
- DPN in UW context
  - Envision joint service with Libraries and central IT to provide a range of services pertaining to digital curation, asset management and preservation – onramp to DPN and/or partners
  - Two team members have been trained to deposit materials into DPN
  - Finalizing deposit agreement with Legal and Business service divisions

# LDAM

Long-term Digital Asset Management Initiative

# Libraries - Digital Preservation Framework

- **Framework for developing and organizing digital preservation policies for the UW–Madison Libraries (LDAM)**
- **Stewardship and preservation**
  - Requires institutional **commitment**
  - **Sustainable** funding
  - **Accountability** – policy, audit, risk
- **Regulatory and programmatic environment**
  - **Federal and state law**
  - **Copyright, fair use, etc.**
  - **Records retention schedules (federal, state, institutional)**
- **International standards and best practices related to digital preservation**
  - **Centrally planned and managed policy and systems - distributed preservation activity**
    - **Technological and geographic redundancy**
    - **Documenting workflows**, to maintain both the **integrity** and **identity** of the **objects** under **curation**.
    - Systems to meet or exceed campus IT **Security requirements, policies, and procedures**
    - Systems considered **mission-critical infrastructure**, with appropriate levels of monitoring, response, and recoverability.



# LDAM – Rationale for Consideration of Campus Service

- Increasingly, scholarly work persist in digital form only
- Recognized need for enterprise scale service to accept, maintain and recover digital assets
- Numerous units support the creation and maintenance of digital archives
- Value in bringing content creators/providers together to consider services for LDAM

# LDAM Initiative - Administration

## Executive Sponsors

Vice Provost for Libraries and University Librarian

Vice Provost for Information Technology and Chief Information Officer

## Governance

Advisory Committee

Working Group

# LDAM – Deliverables

## Report (forthcoming)

- Determines and documents the capabilities required to support the digital archive needs of the University
- Develop inventory and description of services that currently provide digital archive services.
- Assess the level of maturity and alignment of current services with current and future needs of key campus units and the campus as a whole
- Survey community to determine how people are currently addressing digital archiving needs
- Analyze gap between current services and required capabilities
- Survey and describe external opportunities and how they might fit into our plans ( e.g. DPN, DuraSpace, etc.)
- Draft high-level roadmap, communicating campus current state, and the steps necessary to achieve the desired future state
- Make initial recommendations for projects, governance or other actions necessary to achieve the desired future state and evolve the services over time

# LDAM – Phases

- Planning
- Stakeholder engagements (administrative)
- Peer consultations
- Survey to identify communities of need
- Interviews and Focus Groups
- Qualitative Analysis
- Report (forthcoming)

# Process

- Reviewed key tools and documents in use at peer institutions (e.g. OAIS, Data Curation Profile Toolkit, data management frameworks...)
- Reviewed Libraries Preservation Framework
- Developed guiding principles for the LDAM service (for ourselves and to inform stakeholder engagements)
- Developed criteria to facilitate conversations as to nature of service as distinct from backup, storage, etc.
- Environmental scan of existing campus services and technologies

# LDAM Service - Guiding Principles - 1

- Principles for the **creation and operation** of Long Term Digital Asset Management services
  - The Open Archival Information System (OAIS) Reference Model
  - The services(s) will be operated with a focus on the needs of the communities that:
    - Create and wish to preserve content. **OAIS Producers**
    - May wish to access the content. **OAIS Consumers**

# LDAM Service - Guiding Principles - 2

- Responsibility for the **Management** of the Service will be clearly defined...
  - Provide Service **charter** and **scope**.
  - Provide **funding** for the Service – determine pricing models
  - Evaluates Service performance and utilization against defined **goals**.
  - Assesses and addresses **risks** to the Service and holdings.
  - Determine or endorses **policies** related to the Service including:
    - Type of **content** the Service may not, may, is required to accept.
  - **Resolves conflict** involving Producers, Consumers and Service Administration.
  - **Develops and enforces policies for use of the Service**. E.g. must, should, may persist content with the Service.

# LDAM Service - Guiding Principles - 3

**Responsibility** for the **Administration** of the service(s) will be clearly defined and should include.

- Develop and document **policies and procedures** which **ensure**:
    - That the **content is preserved** against the contingencies specified by Management, including demise of the Service.
    - **Conformance** with all legal requirements (e.g. copyright) that apply to the content.
  - Development of formal **succession and escrow** arrangements if the Service discontinues operation.
  - Non-archive **software** should not be relied upon to maintain Representation Data.
  - **Content will never be deleted** unless as part of a **defined strategy** approved by Management. **No ad hoc deletions.**
-



# LDAM Service – Criteria and Considerations for Inclusion

## Balancing cost, risk, and effort

- Metadata - level of description
- Format standardization
- Volatility
- Uniqueness
- Value of content to institution
- Quantity
- Provider profile (commitment to institution)
- Institutional investment and risk

# LDAM - Inventory of Current Campus Services/ Technologies

Library and University Archives repositories: [MINDS@UW](#), [UW Digital Collections](#), [HathiTrust](#)

Shared campus services

[Electronic Lab Notebooks](#) - Lab Notebook Software and Storage for Researchers

[Research Data Services](#) - Data Management Plans, Consultations, Education and Training

[Hitachi Content Platform](#) - DoIT and Library Data Archive Service

[Advanced Computing Initiative](#) - Campus Research Computing Resources

[Campus Computing Infrastructure \(CCI\)](#) - Physical and Virtual Server Hosting and Data Storage

[Bucky Backup](#) - Bucky Backup Lite (Disk based) and Bucky Backup Archive (Tape based)

[UW Cloud Services](#) - Amazon Web Services, Microsoft Azure, Google Cloud Platform, IBM SoftLayer

UW Cloud File Storage – [Box](#), [Google Apps](#), [Office 365](#)

Departmental IT Services – [Social Science Computing Cooperative \(SSCC\)](#),

[Institute for Clinical and Translational Research \(ICTR\)](#), [DiscoverIT Storage](#), [Computer Systems Lab](#),

[Department of Physics](#), [Biochemistry](#), [Bock Labs](#), [Recovering Analog and Digital Data \(RADD\)](#)



# LDAM – Campus Communications and Connections

- Governance Engagement
  - Shared charge with campus governance groups
  - Met with governance groups to review charge and request help promoting initiative
- Survey sent to campus colleges, schools, divisions, and research centers – seeking discussants to inform potential service needs/requirements
- In-depth interviews

# LDAM - Face-to-face Interview Roadmap

1. **Please tell us** a bit more about yourself, your role, your organization (10 minutes)
2. **Please tell us more** about the **digital materials** you are considering preserving. (15minutes)
  - What **value** does the material have that makes it a candidate for preservation?
  - **What aspects** of the materials, information about the material, should be preserved?
3. **In your mind**, what constitutes preservation? (20 minutes)
  - How long?
  - How many copies, where?
  - How accessible should the preserved data be? To whom?\
  - How important is it that the materials be migrated over time to keep up with changes in formats and tools?
  - What are some of the potential uses you see for this preserved data, where some person, group, organization or robot in the future may want to recover these materials for some purpose?
4. **Tell us more** about any regulations, policies, guidelines, mandates, etc. that **influence your preservation decisions and activities**. (10 minutes)
5. Preservation has costs. **Tell us how you think** about the resources (people, storage, etc.) required to prepare materials for preservation, store them, and manage them over time. (10minutes)
  - How much does cost (in terms of money, time, and effort) influence your preservation decisions?
  - Would a campus-wide preservation service be of use to you? If so, what might that look like?
  - Assuming that such a service is unlikely to be 100% centrally-funded, what kind of cost models for such a service would work for you?

# LDAM – Findings - 1

- We interviewed community members to learn more about their units' concerns and approaches to digital preservation at UW
- We conducted in-depth interviews with representatives from eight different units from across campus, representing archival, library, administrative and research perspectives
- No major surprises
  - Several expressed desire for **openness** of materials for **philosophical** and/or **compliance** reasons
  - A few expressed concerns about how best to preserve material that could/should not be open for reasons of **privacy**, **copyright**, or even held elsewhere

'eady

# LDAM – Findings - 2

- Storage and backup solutions **varied** widely
  - Centrally-hosted storage
  - Shared local drives
  - Collaborative cloud platforms such as Box
  - External platforms such as the Open Science Framework
- Consensus that range of **options often felt incomplete** and lacked **formalized workflow**
- Concerns that **time** spent overseeing the technology needs of their unit took away from other job responsibilities
- All aware their LDAM **strategies did not equate to preservation yet** all sought to ensure that their materials were accessible in **perpetuity**

# LDAM – Findings - 3

- Scale of material requiring LDAM across these units
  - Several units cited overall data collections between 1-5TB
  - One unit outside library cited holdings of 90TB
  - One unit generates 50-100TB per year
- Submission preferences – annual submission to daily submissions
- **Takeaway** - Data is being preserved while metadata may not be. Metadata, such as format definitions, is not being saved, so the meaning of the data may be lost, or the data may not be discoverable.
- Continued need to study full range of campus need, particularly for the preservation cases where cultural and campus heritage is at risk

# LDAM Next Steps

- Develop initial LDAM service recommendations and complete draft of report
  - Share report with sponsors and review recommendations
  - Create initial budget model and service layers for local LDAM service if service is endorsed
  - Add HCP to local suite of LDAM services if promising
  - Promote DPN as part of suite of LDAM services endorsed by campus
-



# LDAM - Key Challenges to Campus-level LDAM

- Sustainable and common-good funding models and governance models
- Definitions: clarifying distinctions- storage, archiving, dark archiving, preservation (as set of actions, people, tools)
- Risk analysis vs 'lesser' alternatives
- Copyright and succession of material deposited into DPN or third-party systems

# Next steps with DPN and others

- Support and collaborate with others in the work of the digital library and preservation communities
- DPN
  - Utilize the credits we have earned as a partner – public domain materials only
  - Finalize arrangements with campus legal/business for deposition of additional materials into DPN
  - Formally position DPN as part of campus preservation strategy and add into suite of services
- Libraries will likely be the point of entry to DPN and LDAM services for campus stakeholders
  - Testing and initial ingests via DuraSpace
  - Develop the workflows and best practices that would be adopted by a campus solution in the future