

Michigan's Repository Technology Plans

John Weise

Head of Digital Library Applications, Library Information Technology, University of Michigan



Who are we?

- Library Information Technology [**63 human beings**]
 - Design & Discovery [12 FTE + interns]
 - Digital Content & Collections [15 FTE + intern + part time]
 - Digital Library Applications [14 FTE]
 - Automation, Indexing & Metadata [10 FTE]
 - Architecture & Engineering [10 FTE]
 - AUL/Director and assistant [2 FTE]

What's our approach?

- In our capacity as a **partner**, we enable the divisions of the library to achieve their goals by providing and sustaining technology solutions.
- In our capacity as a **steward**, we seek to unite the preservation, access, and publishing of digital content.

Digital Repository **Services**

Strategically, the desired future is, “**Coherence among the various digital repository and content services through alignment of user experience, policies, and practices.**”

— Library Technology Strategy: Digital Repository Services

Digital Repository & Publishing **Technologies**

Strategically, the desired future is, “**Fundamental digital preservation needs are met by a set of highly reliable repository and publishing technologies that match our principles and policies.**”

— Library Technology Strategy: Digital Repository & Publishing Technologies

Current Repositories

Repository	What	Technology	Service Owner	Size (TB)	Start Year
Digital Collections*	text, image, finding aids, newspapers, other	DLXS, BlackLight, other	LIT	22	1998
Deep Blue Docs	institutional repo	DSpace	Publishing	9	2006
Deep Blue Data	research data repo	Hyrax	Publishing	12	2016
Fulcrum	preservation publishing	Hyrax + Heliotrope	Publishing	1	2017
Dark Blue	Dark repo for audio/moving image, born digital, etc.	Chipmunk (local/custom)	LIT	3	2018
HathiTrust**	digitized books	HathiTrust	HathiTrust	767	2008

*See [U-M Digital Collections by the Numbers](#) for more info.

**HathiTrust is a related bit of work we do, but not directly part of the plans presented herein.

Digital Repository Services Steering Committee

The Digital Repository Services Steering Committee will help **shape, scope and shepherd** the continued work of **all the digital repository services into the future**, collaboratively where it makes sense to do so and separately where user needs and circumstances require it.

Commonality and Variety

Success hinges on **leveraging commonality** in order to achieve sustainable solutions and operational efficiencies.

Success also hinges on **supporting variety** to the greatest extent possible without implementing one-off customizations that impinge on sustainability and introduce operational nuances.

User needs are changing

The user facing elements of the system need to **support ongoing rapid evolutionary changes in user needs** that will outpace changes in most other aspects of the repository system as a whole.

Lessons learned

- Tight coupling of digital preservation and access resulted in **blurred boundaries between preservation functionality and access functionality.**
- A rigid digital collections structure for storing objects and providing access made it **difficult to build arbitrary collections.**
- Enforcing adherence to standards for similar types of content is incredibly effective for content management and use at scale, **however**, supporting a wide variety of content types is essential in order to make sure all objects have a safe long-term home.

ObjectClass (1 of 2)

- ObjectClass is meant to be all-purpose.
- It should by no means be a feature-for-feature replacement of DLXS.
- We have even older things, and some newer things, to consider too.

quombat!
querulous, quarrelsome,
curious wombat



ObjectClass (2 of 2)

- ObjectClass will **depart from the rigid collection and content-type specific structure** of DLXS.
- It will have the potential to **serve a variety of digital preservation and access needs** and go well beyond replacing DLXS, and beyond the needs of our new Digital Collections Platform.
- We need to make sure we **understand the needs of present day stakeholders and end users**, and design a system to meet those needs.

Major System Components

1. Data Preparation Environment
2. Ingest
3. Object Storage
4. Repository Management
5. Descriptive Metadata Management
6. User Role Management
7. Digital Collections Management
8. Rights Management
9. Discovery & Access
10. Usage Statistics & Analytics

What makes a component?

ObjectClass is the sum of parts that communicate across defined boundaries.

- Components should align in purpose with our principles.
- We want to avoid creating unnecessary dependencies.
- A component should be able to serve a number of repository use cases.
- Relative rate of evolution is a factor, e.g. ingest evolves slowly, access evolves quickly
- Who will use it is a consideration.
- Whether a component is for preservation or access is also a differentiator.

1. Data Preparation Environment

- Tools for preparing data for ingest, including normalization, remediation, and packaging.
- Sufficient processing capacity and storage to do the work.

2. Ingest

Repository	Push	Pull	Batch	Single	Self	Mediated	Many	Big
DLXS Texts		✓	✓				✓	
DLXS Images		✓	✓				✓	✓
Deep Blue Docs	✓	✓	✓	✓	✓	✓	✓	
Deep Blue Data	✓	✓	✓	✓	✓	✓	✓	✓
Fulcrum		✓	✓	✓			✓	
Dark Blue		✓	✓				✓	✓

3. Object Storage

Principles

- Standards-based object storage
- Ability to audit storage
- Storage redundancy
- Content recoverable from the file system
- Controlled content migrations
- Guarantee of comparability of access & preservation copies

4. Repository Management

Features and functions should be related to management and monitoring of:

- Administrative metadata
- Technical metadata
- Descriptive metadata
- PREMIS
- Fixity
- Validation
- Access copies
- Access control
- Embargo
- Download of masters
- Migration

5. Descriptive Metadata Management

Repository	Created by	Managed in	Preserved in	Accessed in	Stored with object	Originated in repo
DLXS Texts	collection managers, vendors, publishers	Catalog	XML	XPat	yes	no
DLXS Images	collection managers, vendors, publishers	MySQL	MySQL	MySQL	no	no
Deep Blue Docs	depositor, source	DSpace, Postgres	DSpace, Postgres	Solr	yes	sometimes
Deep Blue Data	depositor, source	Samvera, Fedora	Fedora	Solr	yes	sometimes
Fulcrum	depositor, source	Samvera, Fedora	Fedora	Solr	yes	sometimes
Dark Blue	depositor, source	file system	file system	file system	yes	no

6. User Role Management

This is strictly about defining types of users and what kinds of privileges they have.

7. Digital Collections Management

- Powerful tools for staff in our Digital Content and Collections department who have responsibilities for creating and maintaining digital collections
- Plus, empowered content providers
 - Surging number and scale of partner projects
 - Many types of content
 - Increasingly advanced digital preservation skills and experience
 - Willingness to share the workload

8. Rights Management

- Rights management is complex and by making this a component we endeavor to create tools that help simplify and clarify the process.
- We need methods for documenting the status of every object in ways that are meaningful to repository managers and serve to inform end users.
- This will help us to ensure our rights policies and access controls are in alignment.

9. Discovery and Access (1 of 2)

Rapidly evolving high quality accessible user interfaces

Access technology changes quickly, as do the expectations of users. We'd like all users to have consistently good experiences.

- APIs supporting internal and external uses
- Accessible web applications for public access
- Indexing of metadata and content
- Support for discovery, including integration at the library level

9. Discovery and Access (2/2)

Creating a User Interface for the Digital Collections Platform (parallel project)

The goal is to understand user needs through a needs assessment, competitive analysis, and other research methods.

This is multiple phases of work, being done with the intent of creating a user interface for the Digital Collections Platform that dovetails with the development of ObjectClass.

10. Usage Statistics and Analytics

- We need better usage statistics tools.
- We will leverage data to better understand knowledge-seeking patterns, improve user success in our discovery-to-delivery pipeline, improve services, and contribute to campus learning analytics research and analytics-based decision-making. — Library Technology Strategy: Data Analytics

Next Steps

- Round out the components at a high level
- Create functional requirements for each component
- Investigate applicable technologies
- Plan the roadmap for the Digital Collections Platform and ObjectClass
- Think about the life cycle of the initiative and how to communicate about it
- Continue to share our plans and seek out opportunities for collaboration

Acknowledgements

Noah Botimer
Aaron Elkiss
Roger Espinosa
Kat Hagedorn
Lauren Havens
Nabeela Jaffer
Gordon Leacock
Rob McIntyre
Jeremy Morse
Chris Powell
Esty Thomas
Maurice York

Thank you!

Please feel free to reach out with questions or comments.

John Weise
jweise@umich.edu