

University of Michigan Library Digital Collections by the Numbers

Samvera Virtual Connect - April 24, 2019



whoami

- John Weise
- Head, Digital Library Applications, Library Information Technology
- University of Michigan
- Librarian/programmer turned manager 10 years ago
- Samvera Steering Group member since 2018



What this presentation is about

- Overview of our situation at Michigan
- The good and the not so good of DLXS
- Getting out
- The Top 4 Things

Why?

To open up about our challenges and find out who else is facing the same.



Digital Collections Legacy System

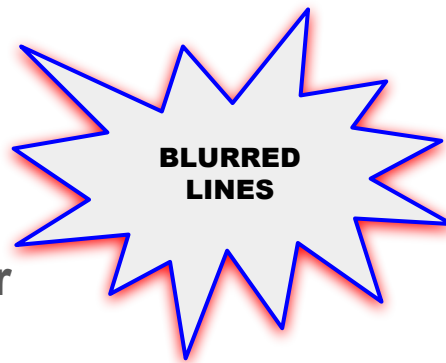
DLXS (Digital Library eXtension Service)

- Access middleware for digital collections
- Digital preservation philosophy underpinning
- 20 or so years... and counting
- Perl has proven to be pretty durable, fwiw



Gearred for Collections

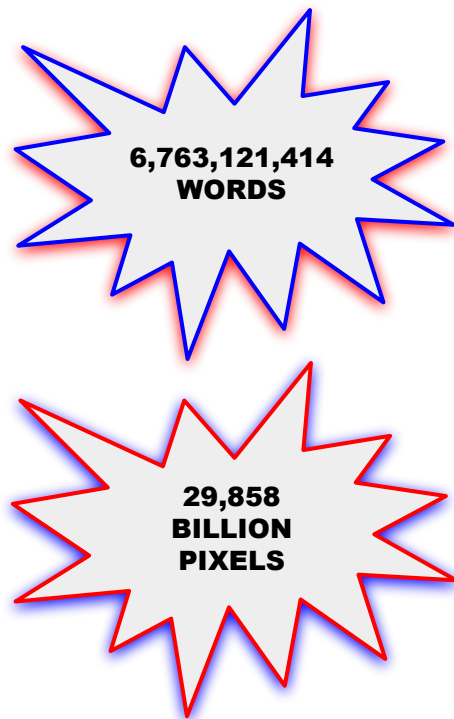
- Heavily collection based
- Tightly coupled preservation and access
 - Configurable at many levels
 - All text/image collections
 - Arbitrary groups of collections
 - Individual collection
- Divided by format and behavior
 - books → go in the text system
 - images → go in the image system
 - manuscripts → **go in one or the other**



Digital Collections in DLXS

- 290+ digital collections, primarily...
 - **Text Collections** in DLXS TextClass
 - Total Collections = 175
 - Total Titles = 159,272
 - Total Size = 4TB
 - **Image Collections** in DLXS ImageClass
 - Total collections = 110
 - Total Images = 2,744,900
 - Total Size = 19TB

<https://quod.lib.umich.edu/t/text>
<http://images.umd.umich.edu/>



Repositories at Play

Repository	What	Now	Next	Year Deployed	Size (TB)	Age (Years)
DLXS Texts	text collections	DLXS TextClass	Samvera	1998	.04	21
DLXS Images	image collections	DLXS ImageClass	Samvera	1999	16	20
Deep Blue Docs	institutional repository	DSpace	Hyrax	2006	9	13
Deep Blue Data	data repository	Hyrax		2016	12	3
Fulcrum	preservation publishing platform	Hyrax + Heliotrope		2017	1	4
Dark Blue	Dark repo for audio/moving image, born digital, etc.	Chipmunk		2018	3	2



What I Like about DLXS

- Scales for production by normalizing content, metadata, workflows
- Highly configurable per collection
- A lot is possible **without touching code**

collections with at least one XSL override	204
collections with at least one CSS override	148
collection level Perl overrides (# of .pm files)	101



It's very difficult at this point to make improvements to satisfy needs and requirements of modern users, especially without a lot of help from developers.




Great support for disparate metadata in ImageClass

Kelsey Museum of Archaeology

Help Search Portfolios Login

Karanis

Download ▾ Share/Cite ▾ Related Views ▾ Problems/comments? New Window



Object Identifier
0000.00.7605

Object Name
Mallet, Small

Lot
University of Michigan excavation (Karanis)

Collections
Kelsey Permanent Artifact Collections

Object Type
Hammers/Mallets

Materials
Wood

Length
12.5 cm

Width
7.0 cm

Height
4.1 cm

Inscribed
no

Full Artifact Description
Small wooden mallet. Wood soft and head surface with decay. Handle broken off and loose in hole.

Verbatim Date
1st c. BC - 4th c. AD

Unit Grid Number
0109

COLLECTION ACCESS
166263 Images/descriptions are openly available.

RIGHTS / PERMISSIONS
The University of Michigan Library provides access to these materials for educational and research purposes. These materials may be protected by copyright. If you decide to use any of these materials, you are responsible for making your own legal assessment and securing any necessary permission. If you have questions about the collection, please contact the Kelsey Museum of Archaeology. If you have concerns about the inclusion of an item in this collection, please contact Library Information Technology.

ADD TO PORTFOLIO
Add

IMAGE SIZE
2272 x 1704

RECORD
0000.00.7605

COLLECTION

Custom metadata is easily configurable in DLXS ImageClass for loading, searching, sorting, and display.

Fields can be mapped to common/core schemas through simple configuration. Changes take immediate effect without reloading, reindexing, or redeploying.



Searching across collections within a class

UMMZ, Museum of Zoology [Help](#) [Search Group](#) [Portfolios](#) [Login](#)

Search Group

You can search all the databases in this group, or use the list of databases to restrict your search to a subset.

To search, enter a word or phrase in a box below and select a field from the menu.
It is not necessary to use more than one box.

Genus	Contains all of these words	
And	Anywhere in record	Contains all of these words
And	Anywhere in record	Contains all of these words
And	Anywhere in record	Contains all of these words

Limit results to:

☐ has digital media

University of Michigan - Museum of Zoology Databases

Select one or more databases to search. Click the database name to search it alone, often in more detail.

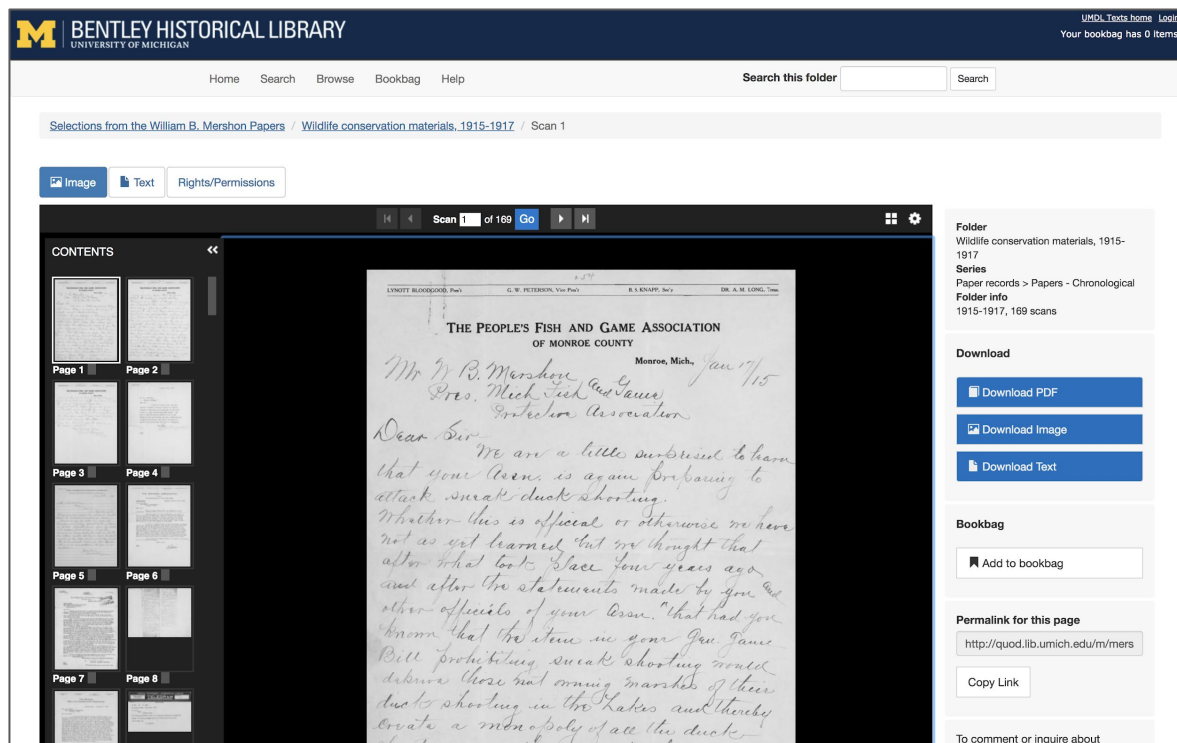
- ☒ University of Michigan Museum of Zoology, Amphibian and Reptile Audio Files
- ☒ University of Michigan Museum of Zoology, Amphibian and Reptile Catalogue
- ☒ University of Michigan Museum of Zoology, Amphibian and Reptile Type Specimens
- ☒ University of Michigan Museum of Zoology, Distribution Maps of Michigan Fishes
- ☒ University of Michigan Museum of

Image collections mapped to a common metadata schema can be searched together by that metadata schema.

Searching a single collection, all of its special fields are available.



IIIF support and Universal Viewer integration



<http://quod.lib.umich.edu/m/mershon/851844.0026.017/1>

John Weise - University of Michigan Library Digital Collections by the Numbers



Search results with thumbnails and highlighting

The screenshot shows the Bentley Historical Library website interface. At the top, the header includes the University of Michigan logo, the text 'BENTLEY HISTORICAL LIBRARY UNIVERSITY OF MICHIGAN', and links for 'UMDL', 'Texts home', and 'Login'. A notification states 'Your bookbag has 0 items'. Below the header is a navigation bar with 'Home', 'Search', 'Browse', 'Bookbag', and 'Help'. The main content area shows the search path 'Joseph L. Hudson Papers / J.L. Hudson Company, 1967 / Folder Level Search Results'. A blue box indicates the search results: 'You searched this collection for **detroit** in full text, which was found 4348 times in 29 folders.' Below this are buttons for 'Back to search' and 'Refine this search'. A pagination bar shows 'Folder 1 - 6' and '6 folders'. A 'Sort by:' dropdown is also present. The search results are displayed in a list. The first result, 'Scan 1', shows a thumbnail of a document page. To the right of the thumbnail, the text reads: 'Folder J.L. Hudson Company, 1967', 'Series Other Materials', and 'Folder info 1967, 3 scans'. Below this, a snippet of text is shown: '... that you have declared an intention to stop patronizing Hudson's because of the actions of the NEW **DETROIT** committee. You are, of course, entirely free to make that decision if you wish, but I hope y...'. The second result, 'Scan 2', follows a similar format with its own thumbnail and metadata.



<https://quod.lib.umich.edu/h/hudson?didno=85215.0001.030:rgn=full+text:size=25:sort=freq:start=1:subview=detail:type=simple:view=reslist:q1=detroit>

John Weise - University of Michigan Library Digital Collections by the Numbers

Advanced support for structured text in TextClass

The screenshot shows the EEBO Text Creation Partnership website. At the top, there's a navigation bar with 'Home', 'Search', 'Browse', 'Bookbag', and 'Help'. The 'Search' tab is active. Below it, the 'Proximity Search' section is highlighted. The search interface includes a 'Search in:' dropdown menu with 'drama' selected. Below this, there are two 'Find:' sections. The first section has a text input field, a 'near' dropdown, and a distance dropdown set to '40'. The second section has a 'near' dropdown, a text input field, and a distance dropdown set to '40'. A 'Tip' section provides examples: 'work*' finds 'worker,' 'working,' etc., and 'war and peace' finds 'war and peace'. There's a 'More tips' link. The 'Limit to:' section has two rows, each with a text input field, an 'in' dropdown, and an 'author' dropdown. The 'Restrict to Navigations series:' dropdown is set to 'No'. The 'Restrict by date range:' section has two date dropdowns set to '1470' and '1700', with an 'and' operator between them. A 'Search' button is at the bottom. The footer mentions 'Powered by DLXS' and provides contact information for content and errors.

Early English Books Online
Text Creation Partnership

UMDL Texts home Login
Your bookbag has 0 items

Home Search Browse Bookbag Help

Proximity Search

Basic Boolean Proximity Search Bibliographic Word Index History

Search in: drama

Find: ([] near [40])

[near] [] [40]

Tip: work* finds "worker," "working," etc.
war and peace finds "war and peace"

▼ More tips

Limit to: [] in [author]
and
[] in [author]

Restrict to Navigations series: [No]

Restrict by date range: [1470] and [1700]

Search

Powered by DLXS
To comment or inquire about content, contact eebotcp-info@umich.edu
To report errors, contact [DLPS Help](#)

The ability to search full-text within specific regions of structured XML is a strength of TextClass.

Results can also be limited by metadata.

It's more power than most users need or want, but for some it is just the thing.



Top 10 Google searches for our digital collections

- trust in the lord with all your heart
- **middle english dictionary**
- proverbs 3 5-6
- love your neighbor as yourself
- 10 commandments kjv
- judge not lest ye be judged
- the lord's prayer kjv
- proverbs 3:5-6
- love thy neighbor as thyself
- the spirit of the lord is upon me



Shiny New Middle English Dictionary

 Middle English Compendium

About Help Contact Us

Dictionary Bibliography Quotations

Middle English Dictionary

The world's largest searchable database of Middle English lexicon and usage for the period 1100-1500. An invaluable resource for lexicographers, language scholars, and all scholars in medieval studies.

[Read more about the dictionary](#)

Image: The Ellesmer Manuscript of Geoffrey Chaucer's The Canterbury Tales, c. 1400-1405. Shown is the introduction to "The Knight's Tale." Available online at the Huntington Library.

How to get started

If you want to do more than just look up a word, learn about Dictionary,

More ways to search

You can change the scope of your search through the use of boolean

History of the MED

Almost 90 years in the making, the MED will always be a team effort and

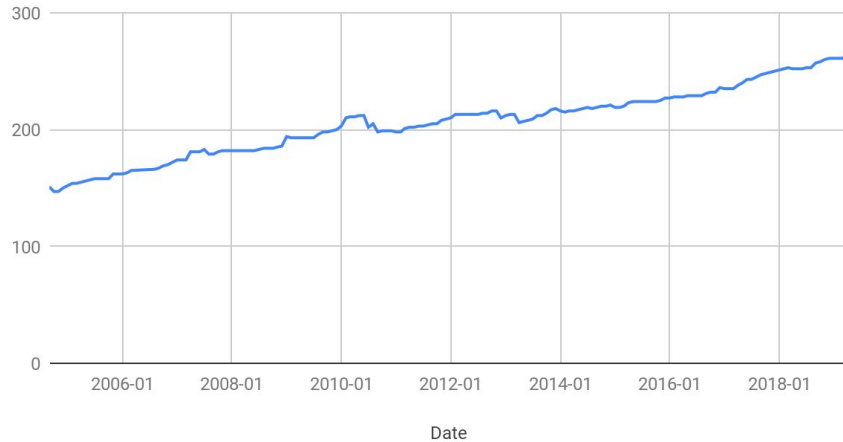


<https://quod.lib.umich.edu/m/middle-english-dictionary/dictionary>

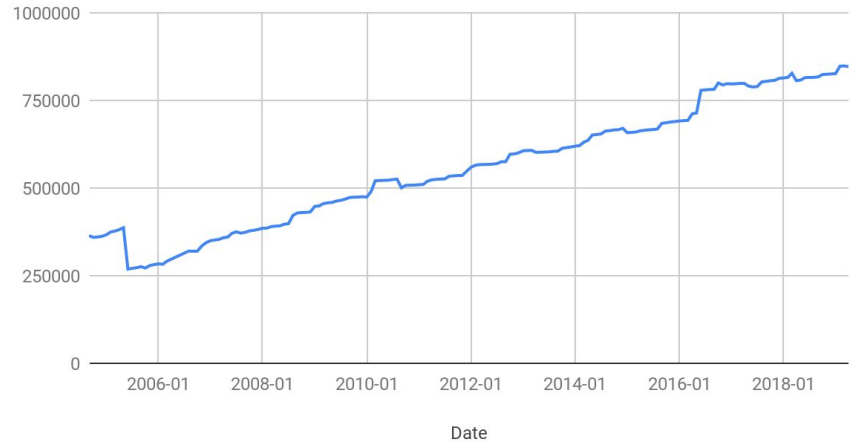
John Weise - University of Michigan Library Digital Collections by the Numbers

Growth of collections

Number of Collections



Objects



Collections created or updated in last year

Text Collections

3 created & 12 updated at least once



Image Collections

5 created & 21 updated at least once



Content Sources

Text Collections

Michigan Publishing	80
Humanities Text Initiative	33
Licensed (by U-M Library)	23
Digital Library Production Service	7
U-M Library	6
Bentley Historical Library	6
Clements Library	5
Text Creation Partnership	4
U-M Dentistry Library	3
University of Michigan	2
University of Michigan Herbarium	1
Taubman Library	1
State of Michigan	1
Special Collections Research Center	1
English Language Institute	1
Abraham Lincoln Association Publications	1
	175



Partnerships! Relationships!
We provide services to humans!

Image Collections

Special Collections Research Center	18
University of Michigan Museum of Zoology	14
Bentley Historical Library	12
University of Michigan Herbarium	8
Michigan Publishing	8
U-M Library	5
History of Art Department	5
University of Michigan	3
U-M Art Architecture & Engineering Library	3
Saline District Library	3
Licensed (by U-M Library)	3
Learning Resource Center	3
Kelsey Museum of Archaeology	3
Digital Library Production Service	3
Asia Library	3
Ann Arbor District Library	3
U-M Dearborn	2
Museum of Anthropological Archaeology	2
Art Images for College Teaching	2
Ypsilanti Historical Society	1
University of Michigan Museum of Art	1
U-M Clark Library	1
School of Music and Dance	1
James and Anne Duderstadt	1
College of Pharmacy	1
Clements Library	1
	110



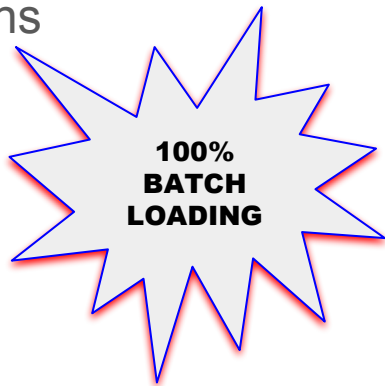
How do collections get into DLXS?

Many different...

- routes
- formats
- transformations
- configurations
- remediations
- repetitions

Always...

- mediated



Marvelous Museums!

The College of Literature, Sciences, and the Arts is the home to all campus museums.

They have consolidated all collection management databases on to a common platform giving us a single point of contact for technology automation and workflow integration.



Who does the work?

3 service staff in Digital Content & Collections
(DCC) primarily deploy collections

Text Collections – Chris Powell

Image Collections – Rob McIntyre

Project Manager – Lauren Havens

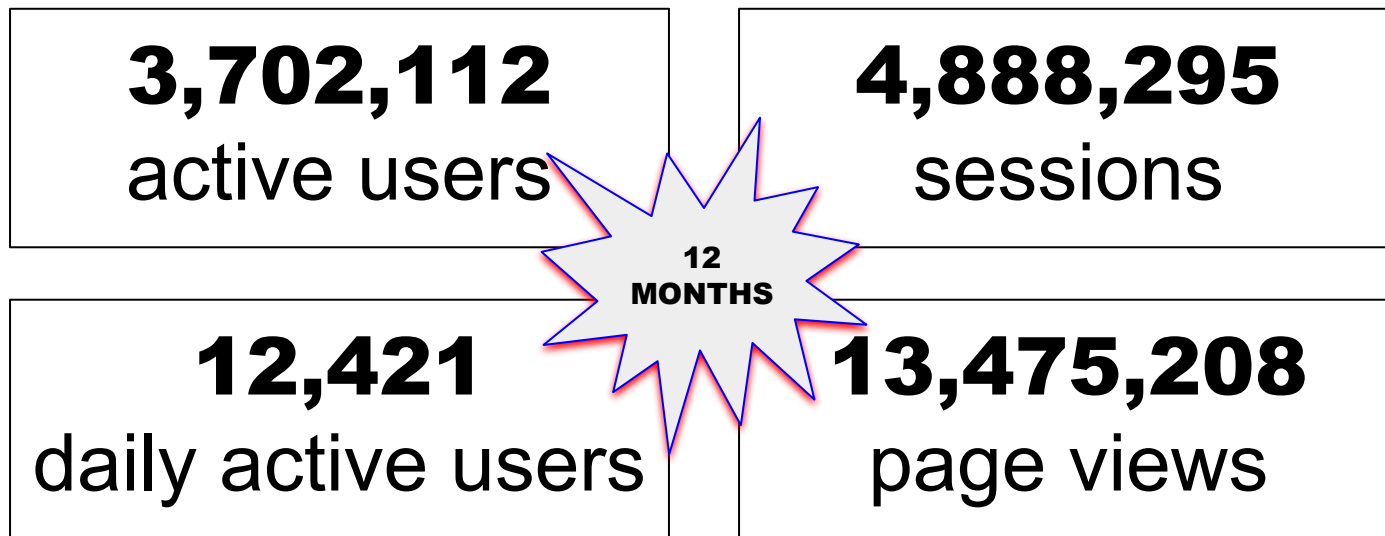
Numerous others play important supporting roles.
Michigan Publishing does their own.

Library IT has **55** staff in **4** departments

- Architecture & Engineering
- Digital Content & Collections
- Design & Discovery
- Digital Library Applications

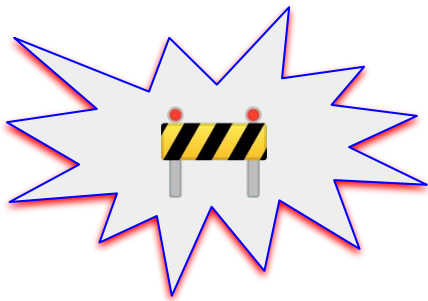


How much are the collections used?

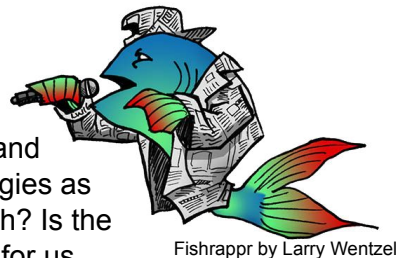


ObjectClass — The DLXS Ouster

- No more silos of content dependent upon silos of code
- Store objects of any kind
- Include any object in any collection(s), or in no collection at all



ObjectClass isn't a snazzy name like [Cozy Sun Bear](#), [Heliotrope](#), [Chimera](#), and [Fishrappr](#). Ironically, it suffers from the same ambiguous mixing of terminologies as TextClass and ImageClass. Is it a class of content, or a class of code, or both? Is the object code, content, or both? Even so, ObjectClass is the name that works for us.



How will we get everything out of DLXS?

- Objects are stored in file system
- However, important behaviors are buried in code customizations



- Collection level configuration information is valuable too



Implicit to Explicit

- Approaching the content from all angles
 - collection level configuration
 - files in file system
 - indexes
 - databases
 - **XML output of DLXS middleware**



We can go anywhere

- Flexibility to go in a number of different directions.
- Setting ourselves up for positive forward motion, with better outcomes, and lower maintenance, for everything we do in the future.

Unambiguous structure, relationships, and metadata that can support a variety of uses.

- Using standards like BagIt, METs and OCFL for storing preservation-complete objects in the file system.



The Top 4 Things



Rock solid core digital preservation functions

- Ingest
 - Validation
 - Fixity
 - All types of content
 - Explicit and unambiguous
 - Redundancy & geographic separation
 - Scalable
 - No bottlenecks
 - Common across our repositories
 - Minimal rigid dependencies on software or infrastructure
 - Policy driven object management
 - Reliable and stable
- We'd like to be really good at getting stuff into the repository, knowing what's there, and keeping it safe.
 - Efficient
 - Accountable



Flexible metadata support

- Many sources
- Many schemas
- Mapping/crosswalking
- Batch/bulk
- Asynchronous from object ingest
- Manual edits, logged and reported
- Analysis/reporting

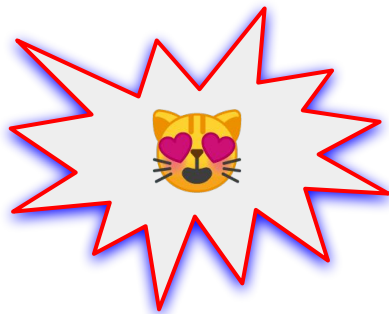
Metadata is disparate and flows at a pace different from the objects. We want to be flexible and accommodating while also encouraging good practice.



3 **Rapidly evolving** high quality user interfaces

- Search
- Discovery
- Access
- Accessibility
- Usability
- Integration
- Authentication
- Authorization
- Rights management
- APIs (IIIF, OAI, ...)
- Sharing
- Re-use

Access technology changes quickly, as do the expectations of users. We'd like to consistently provide users with good experiences getting the stuff they need.



4 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
- Need for repository management applications that allow content providers to see, understand, and manage their objects

Adaptability and scalability of infrastructure, workflows, policies, and human capacity are super important for providing and sustaining digital preservation and access services.

