STRATEGIES AND TOOLS FOR DIGITAL REPOSITORY SELECTION AND MIGRATION

TODD CROCKEN - ANNE WASHINGTON
DIGITAL INITIATIVES SYMPOSIUM 2019

Basics

- 2 Year IMLS grant project (lg-70-17-0217-17)
- Based at University of Houston with partner institutions: University of Victoria, University of Miami, and IUPUI.
- Advisors/Stakeholders: Stanford, Indiana, DPLA, Duraspace.

Intros

Todd Crocken (he/him/his)

- Content Strategist for Bridge2Hyku
- 10 years of experience doing content migration outside academia
- Loves Tortas

Anne Washington (she/her/hers)

- Metadata Strategist for Bridge2Hyku
- 3 years on UH digital library migration project, 6 years experience with Samvera technologies
- Loves pizza

You? (Name, Job, Favorite Food)

Schedule

- System Evaluation and Selection ~ 60 mins
  - Short break ~10
- Migration Strategy ~ 60 mins
  - Short break ~10
- Migration Resources ~ 30 mins

Pairs, Groups, or ...?

- Lets self-select groups of mixed skills.

Goals

- Define evaluation criteria of your new platform by creating use cases and user stories.
- Identify your migration's possible complexities through planning a roadmap
- Identify resources and tools for migration

Why Migrate?

- Cultural?
  -
  -
- Technical?
  -
  -
- Economic?
  -
  -

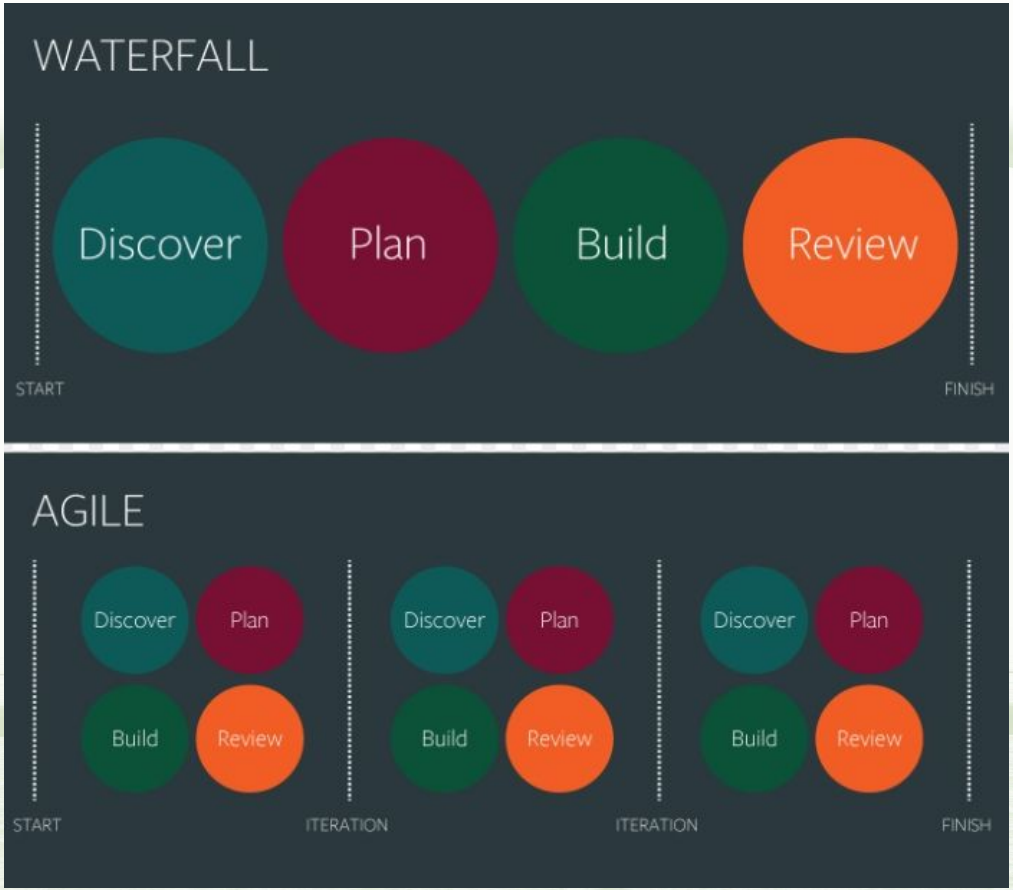
Why Migrate?

- Cultural?
  - New emphasis on digitized collections
  - Digital Humanities
- Technical?
  - Metadata wrangling
  - Open source
- Economic?
  - Fees vs Skills
  - AWS

Why Migrate? – Rationales for UHL

- CONTENTdm sunsetted its local version in 2017.
- We needed a more robust DAMS for new digital initiatives
  - Is more flexible, scalable, interoperable
  - Manages larger amount of data in a variety of formats
  - Accommodates creative workflows
  - Allows for configuration of additional functionalities
  - Supports and enables linked data

Approaches to System Selection and Evaluation



https://blog.ganttpro.com/en/waterfall-vs-agile-with-advantages-and-disadvantages

"WATERFALL"

UH Evaluation & Selection Process

- Conducted needs assessment
  - Stakeholder requirements
    - Internal focus groups
  - Identified future needs
    - Reviewed aspirational institutions
    - Reviewed scholarly literature
  - Gathered data
    - Identified materials and users

"WATERFALL"

UH Evaluation & Selection Process

- Developed selection criteria based on needs
  - System Function
  - Content Management
  - Metadata
  - UI/Search
- Established evaluation methods

"WATERFALL"

UH Broad Evaluation Methodology

Scoring:

- Scores were determined by reviewing supporting documentation, marketing information, or talks with vendors
- 0 - Does not support criteria
- 1 - Supports criteria
- Scores were totaled up and top 2 systems were evaluated for the final round

DAMS	Evaluation Score	Possible Score
Fedora	27	29
Fedora / Hydra	26	29
Fedora / Islandora	26	29
Collective Access	24	29
DSpace	24	29
CONTENTdm	20	29
Rosetta	20	29
Trinity (iBase)	19	29
Preservica	16	29
Luna Imaging	15	29
RODA	6	29
Invenio	5	29

"WATERFALL"

UH Detailed Evaluation Criteria

- Drawn from the same sources
- Looked at specific features for DAMS

"WATERFALL"

UH Detailed Evaluation Criteria

- Criteria was divided into eight testing sections
  - System Environment and Testing
  - Administrative Access
  - Content Ingest and Management
  - Metadata
  - Content Access
  - Discoverability
  - Report and Inquiry Capabilities
  - System Support

"WATERFALL"

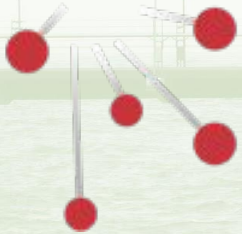
UH Evaluation Criteria

Broad (29)	Detailed (131)
• System Function • Content Management • Metadata • UI/Search	• System Environment and Testing • Administrative Access • Content Ingest and Management • Metadata • Content Access • Discoverability • Report and Inquiry Capabilities • System Support

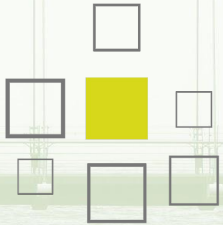
"WATERFALL"

System Setup for Detailed Evaluation

- Virtual servers were set up for each system
- Latest stable versions were chosen (no betas) and installed
- All supporting server software was installed
- Additional software to support the evaluation testing was installed and set up



Fedora™



DSPACE

"WATERFALL"

Gathering Collections for Testing

- Wide variety of item formats currently available
- Variety of item formats for future projects
- Large collections
- Included items with bad metadata, wrong formats, and corrupted data

"WATERFALL"

Detailed Evaluation Methodology

Scoring:

- Each system was tested with the same set of collection items
- Ranged score 0 - 3 (0 failed - 3 fully supported)
- Yes / No criteria was scored 0 or 3
- System documentation was still used in some scoring

Process:

- Scores were tallied to show top ranked systems
- Systems summarized into advantages and disadvantages
- Recommendation was compiled and written into the final report

Score of Top 2 DAMS from UHL Testing Using Detailed Evaluation Criteria

Testing Sections	DSpace Score	Fedora Score	Possible Score
System Environment and Testing	21	21	36
Administrative Access	15	12	18
Content Ingest and Management	59	96	123
Metadata	32	43	51
Content Access	14	18	18
Discoverability	46	84	114
Report and Inquiry Capabilities	6	15	21
System Support	12	11	12
TOTAL SCORE:	**205**	**300**	**393**

Fedora/Hydra - Advantages and Disadvantages

Advantages	Disadvantages
Open source	Steep learning curve
Large development community	Long setup time
Linked Data ready	Requires additional tools for discovery
Modular design through API	No standard model for multi-file objects
Scalable, sustainable, and extensible	
Batch import / export of metadata	
Handles any file format	

DSpace - Advantages and Disadvantages

Advantages	Disadvantages
Open source	Flat file and metadata structure
Easy installation / turnkey system	Limited reporting capabilities
Existing familiarity through TDL	Limited metadata features
User group / profile controls	Does not support Linked Data
Metadata quality module	Limited API
Batch import of objects	Not scalable or extensible
	Poor user interface

"WATERFALL"

Where UHL ended.

- Created a Report
- Left the ETDs in DSpace
- Digital Library
  - Hyku (Hydra-in-a-Box)
  - Hyrax

Do we have 6-12 months in this workshop?

Do we have 6-12 months in this workshop? (no)

"AGILE"

Agile 101

- Software development-focused
- Project management framework
- Cross-functional team
- Incremental, iterative work

"AGILE"

Manifesto for Agile Solution Development

Individuals & interactions	***over***	**Processes & tools**
Working solutions	***over***	**Comprehensive documentation**
Customer collaboration	***over***	**Contract negotiation**
Responding to change	***over***	**Following a plan**

"AGILE"

MoSCoW Prioritisation

- Must Have
- Should Have
- Could Have
- Won't Have this time

"AGILE"

Minimum Viable Product

- What can the stakeholders and developers agree is the feature set required to launch the platform?
- "Bells and whistles" vs the actual car

"AGILE"

What's not viable about your current system?

"AGILE"

Use Cases?

- What are your broader institutional need?
  - ETD / IR?
  - Cultural Heritage Objects?
  - Special Collections?
- What are your unique needs?
  - Lots of data sets?
  - Lots of video?
  - Starting from scratch?

Worksheet time

"AGILE"

User Stories

- Testable
- Much smaller unit of work/development

Basic Rules

- Don't get technical
- Keep it short
- Be clear and avoid using phrases "maybe", "something like", etc.
- Avoid functional requirements, i.e. "click a button"
- Avoid conjunctions AND/OR

"AGILE"

User Stories

- Construction

As a **WHO** I want **WHAT** so that **WHY**.

How the user stories can help after this workshop?

- User stories are still a good way to bring your needs to consortiums and SaaS providers.
- Begin to think about using project management frameworks
- Kickstart asking the more nuanced questions about what your institution really needs from a new platform

BREAK

Migration Planning Worksheet

- Fill out during break, or while we discuss migrations in second part of this workshop
- We're here to answer any questions
- It is ok to only have a vague answer

Between Selection and Migration

- Shouldn’t be a “baton passing”
- Should be the same group that makes the selection
    - But who are those people?

Stakeholders?

- Department Heads
- “Business Development” people
- Special Collections?
- Institutional Publisher?

Team Building

- Subject Matter Experts?
  - Special Collections?
  - Scholarly Communications?
- IT? Dev?
- Metadata Librarians

Timeline

- How long is the time between your selection and implementation?
  - What does “implementation” mean in your context?
    - Ready to migrate
      OR
    - All migration done
  - Do you need to reevaluate anything?
    - MVP!

So you're migrating?

Next up

- Migration Planning
- Normalize Metadata, Modify Content
- Migrate!
- Verify

Migration Planning

- Contextual Considerations
- Digital Library Analysis
- Content Analysis
- Metadata Analysis

Migration Planning

- **Contextual Considerations**
  - Type
  - Size
  - Resources

Migration Planning

- **Contextual Considerations**
  - Audience
  - Administration
  - Dependencies

Migration Planning

- **Digital Library Analysis**
  - Systems
  - Size of collection(s)
  - Data model

Migration Planning

- **Content Analysis**
  - Work types

Migration Planning

- **Content Analysis**

Collection	Objects	Photographic		Document		Complex	
		Single-sided	Double-sided	Single Page	Multi-Page	Hierarchical	Audio + Document
Galveston 1915 Hurricane Photographs	124	X					
Burdette Keeland Architectural Drawings and Photographs	75	X	X		X		X

Migration Planning

- **Content Analysis**
  - File types
  - File locations

Migration Planning

- **Metadata Analysis**
  - Schema use
  - Field requirements
  - Controlled vocabularies

Migration Planning

- **Metadata Analysis**
  - Copyright
  - Data quality
    - Consistency
    - Standards

Discussion

- Migration Planning
  - Contextual Considerations
  - Digital Library Analysis
  - Content Analysis
  - Metadata Analysis

Review and revise content

- Normalize metadata
- File management
- Other?

Review and revise content

- **Normalize metadata**
  - Why
    - Very important component of migration
    - Stakeholder communication
  - What
    - Identify and prioritize issues

Review and revise content

- **Normalize metadata**
  - How
    - Manual
    - Programmatic
  - When
    - Before
    - In transit
    - After

Review and revise content

- **File management**
  - File locations
  - Filenaming
  - Preservation

Discussion

- Review and revise content
  - Normalize metadata
  - File management
  - Other?

Migrate Content

- Map Metadata
- Model Work Types
- Source Repository Export
- Target Repository Import

Migrate Content

- Map Metadata

Your Repository	Target Repository
Title (dc.title)	Title (dc.title)
Description (dc.description)	Abstract (dct.abstract)
Creator (<name><namePart>, etc..)	Creator (dc.creator)

Migrate Content

- Map Metadata

Your Repository	Target Repository
Title (dc.title)	Title (dc.title)
Description (dc.description)	Abstract (dct.abstract)
Creator (<name><namePart>, etc..)	Creator (dc.creator)

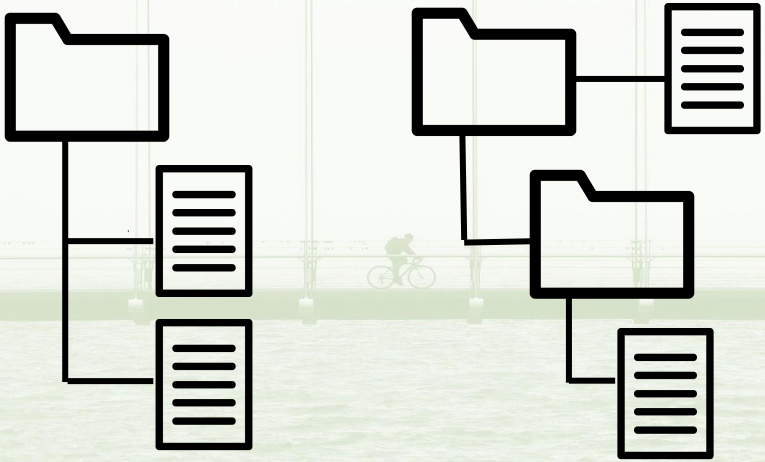
Migrate Content

- Map Metadata

Your Repository	Target Repository
Title (dc.title)	Title (dc.title)
Description (dc.description)	Abstract (dct.abstract)
Creator (<name><namePart>, etc..)	Creator (dc.creator)

Migrate Content

- Model Work Types



Migrate Content

- Source Repository Export
- Target Repository Import

Content Verification

- Approaches
- Measures

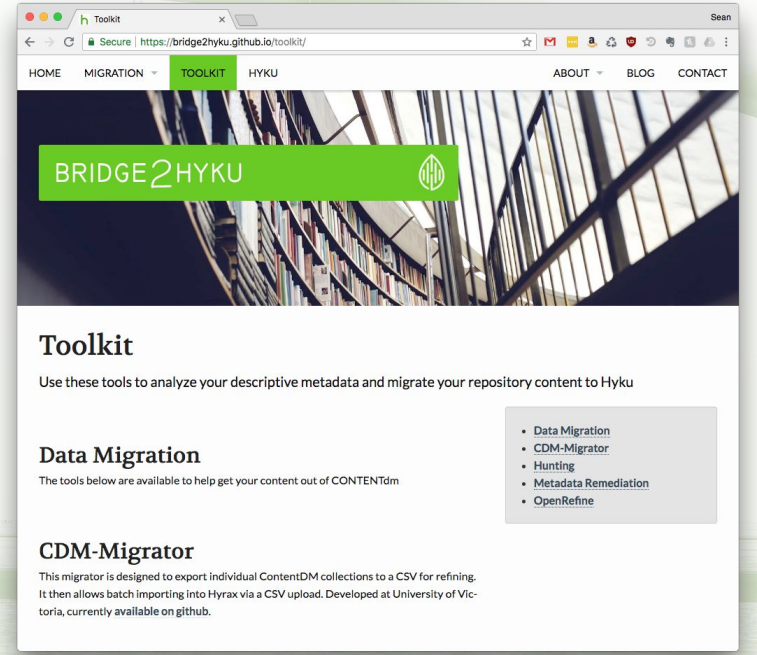
BREAK

Migration Resources

- B2H Toolkit
  - Website
  - CDM Bridge
  - HyBridge
- Academic Articles
- Communities

The Website

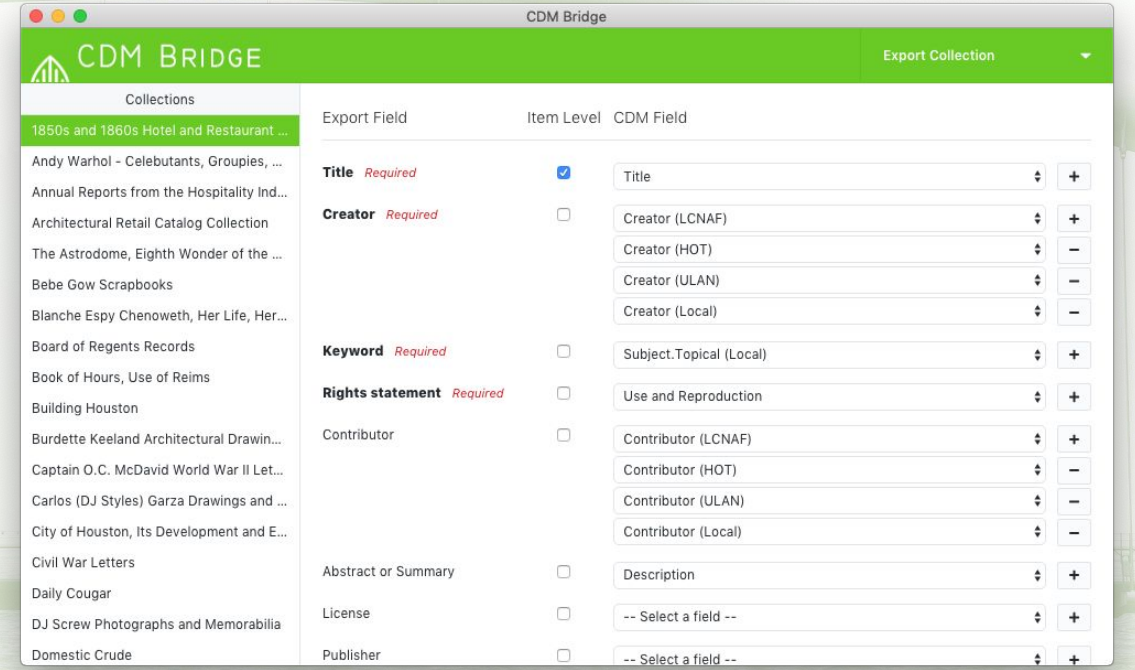
- Migration best practices
- Online resources
  - Blogposts about specific issues
  - Links to other useful sites
  - Migration framework
- Toolkit instructions and info
- End of Phase 1&2 Reports



CDM Bridge

http://bit.ly/cdm-bridge

- Desktop Electron-app
- Works with hosted CDM
- Metadata crosswalking
- Export csv and files



HyBridge- http://bit.ly/hykubridge

- Ruby Gem for Hyrax and Hyku
- Creates easy to use tab for bulk ingest
- Allows collection-specific ingest
- Seamlessly integrated with CDM Bridge's output

Articles- http://bit.ly/migration-articles

- ***Breaking Up With CONTENTdm: Why and How One Institution Took the Leap to Open Source*** - Heather Gilbert and Tyler Mobley (April 2013)
- ***Evaluating, Planning, and Completing a Successful Migration: A Case Study*** - Elizabeth Chance (2018)
- ***Are we still working on this? A meta-retrospective of a digital repository migration in the form of a classic Greek Tragedy (in extreme violation of Aristotelian Unity of Time)*** - Steve Van Tuyl, Josh Gum, Margaret Mellinger, Gregorio Luis Ramirez, Brandon Straley, Ryan Wick, Hui Zhang (August 2018)

Communities

- User Groups
  - DSpace User groups
  - Fedora User groups
  - CONTENTdm User groups
- Slack Servers
  - Samvera
    - #hydra-migration
  - Code4lib
  - Fedora-Project
  - Metadata Support Group

Extras?
Comments?
Thoughts?

CONTRIBUTE

- Join CONTENTdm Migration IG
  on samvera slack: #hydra-migrations
- Have a migration story for a blogpost or want to help with development/documentation?
  Contact Todd Crocken (tcrocken@uh.edu) or
  Andy Weidner (ajweidner@uh.edu)
- Follow us on Twitter! @BRIDGE2Hyku
- Check out our github: https://github.com/Bridge2Hyku

THANKS!

Background from unsplash.com
Abstract: As digital repositories evolve, so to do the needs of institutions who employ them. Increasingly, institutions are faced with the daunting task of migrating content from one repository to another. But what strategies exist to help institutions identify suitable repositories and effectively and efficiently plan and execute a migration? This workshop aims to explore the issues and strategies of repository (re)selection and migration.

Participants will learn about the different phases of a migration process including: system evaluation and selection, migration planning, and implementation strategies and tools. Throughout the workshop, participants will actively explore these phases as they relate to their organizational context and come away with questions and next steps for planning for system selection and/or a migration at their own institution.

The workshop will be led by members of the Bridge2Hyku (B2H) Project, an Institute for Museum and Library Services (IMLS) National Leadership/Project Grant (LG-70-17-0217-17) initiative led by UH Libraries to support the creation of the B2H Toolkit- a suite of resources for migration planning and implementation.