

# Implementing Open Source Systems for Digital Asset Management and Preservation

Andy Weidner, Drew Krewer,  
Bethany Scott, Sean Watkins

UNIVERSITY of **HOUSTON** | LIBRARIES

Texas Conference on Digital Libraries | Austin, TX | May 26, 2016

# Overview

Background  
Preservation System  
Access System  
Next Steps

# Background

*...with Andy and Drew*



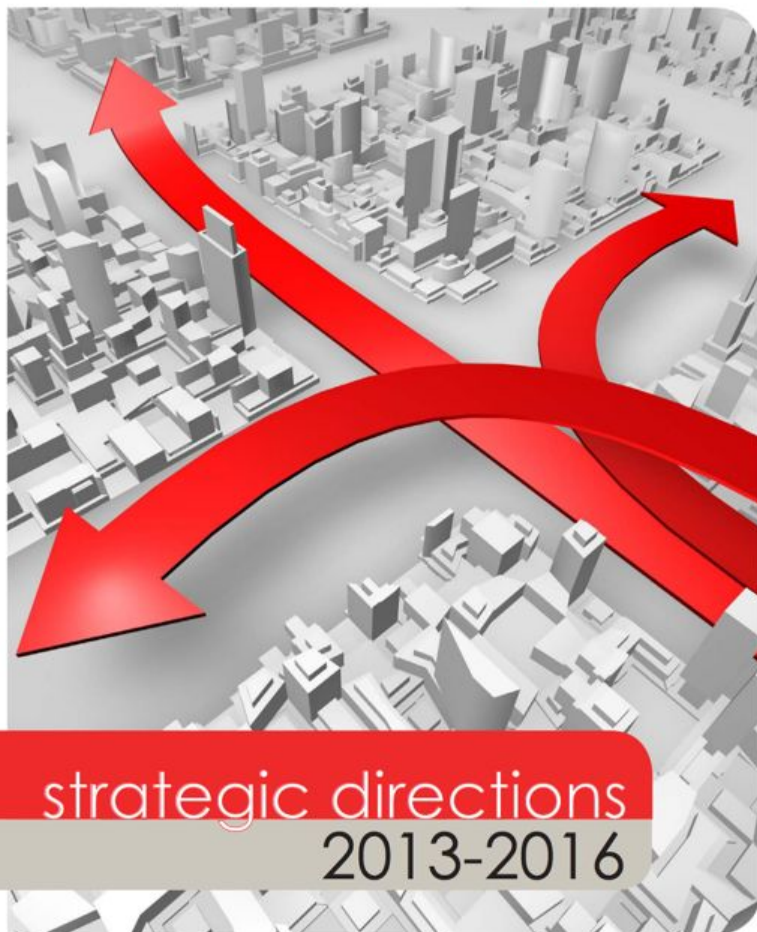
# UH Libraries Task Forces

- 2014 to Present
- Digital Preservation Task Force (DPTF)
- Digital Asset Management System (DAMS) Implementation Task Force



# UH Libraries Strategic Directions

UNIVERSITY of **HOUSTON** | LIBRARIES



“...work assiduously to expand our unique and comprehensive collections that support curricula and spotlight research. We will pursue seamless access and expand digital collections to increase national recognition.”

UNIVERSITY of **HOUSTON** | LIBRARIES

# DAMS Task Force

- Evaluate DAMS Landscape
- Identify Viable Replacement Systems
- Test Systems
- Make Recommendation
- Implement System

# Digital Preservation Task Force

- Define Scope
- Articulate Priorities and Policies
- Determine Resources
- Align with TDL
- Recommend Next Steps

# Digital Access & Preservation Working Group

- Consists of members of the DAMS Task Force & the Digital Preservation Task Force
- Iterative development of workflows, tools, and documentation
- Transparently address preservation concerns throughout development lifecycle
- Communicates progress



# The DAPWG Team

## Metadata & Digitization Services

Digitization  
Services  
Coordinator  
*(Drew Krewer)*

Metadata  
Services  
Coordinator  
*(Andy Weidner)*

Metadata  
Librarian  
*(Anne Gaynor)*

## Web Services

Digital  
Repository  
Developer  
*(Sean Watkins)*

## Special Collections

Digital  
Projects  
Coordinator  
*(Bethany Scott)*

# Preservation System

*...with Drew and Bethany*

UNIVERSITY of **HOUSTON** | LIBRARIES

# Archivematica Features

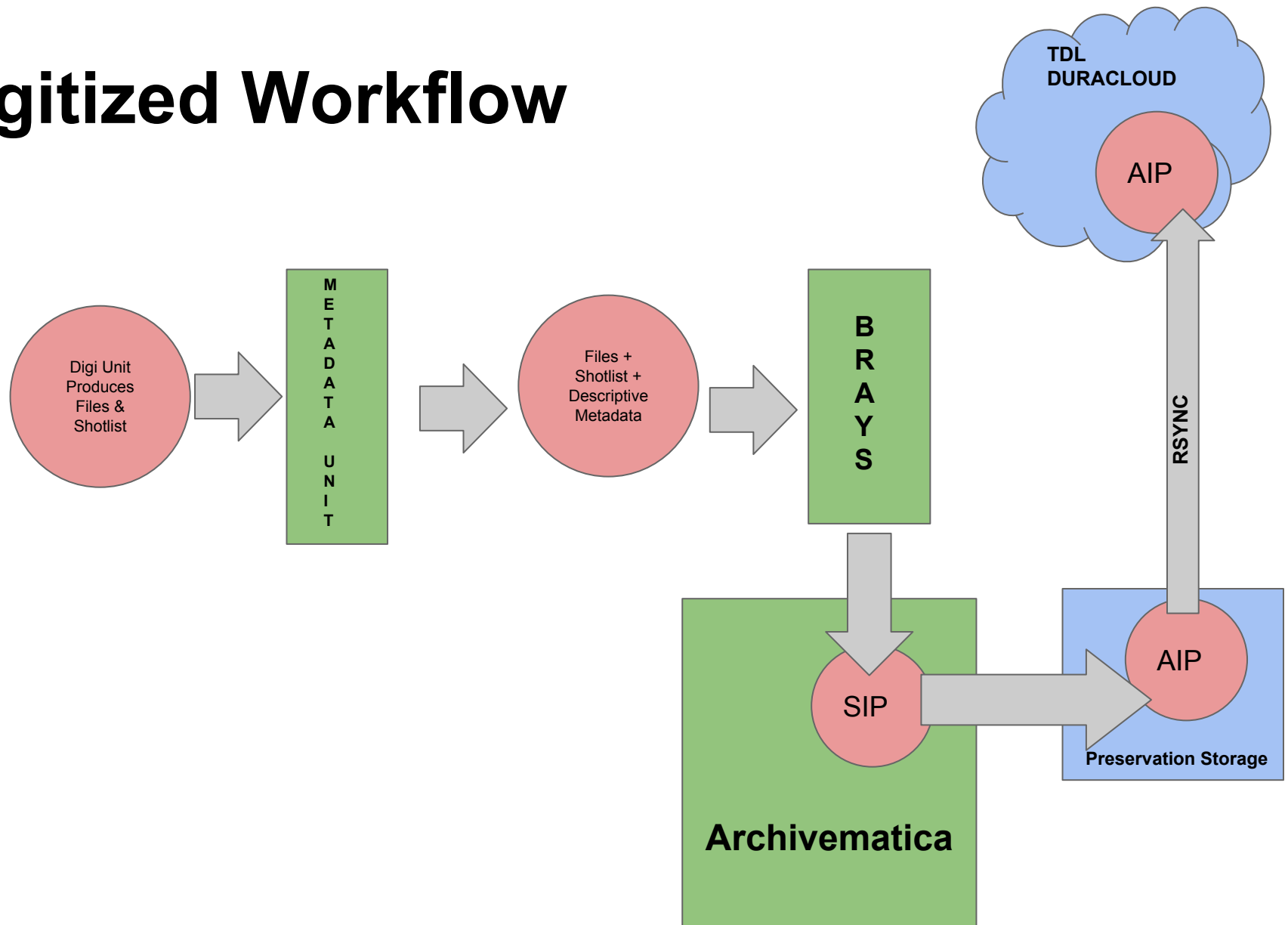
- Based on OAIS model & TDR specifications
- Uses METS, PREMIS, Dublin Core, and the LoC BagIt specification
- Open Source
- Microservices Model
- Customizable



# Born Digital Workflow

- Acquire digital files on removable storage media
- Add collection/item info to digital media accessioning log
- Create disk image using FTK Imager
- Run bulk\_extractor and fiwalk reports in BitCurator
- Create core descriptive metadata in metadata.csv
- Structure package for Archivematica
- Transfer and ingest through AM
- Create digital object in ArchivesSpace and add ARK

# Digitized Workflow



# Access System

*...with Andy and Sean*

UNIVERSITY of **HOUSTON** | LIBRARIES



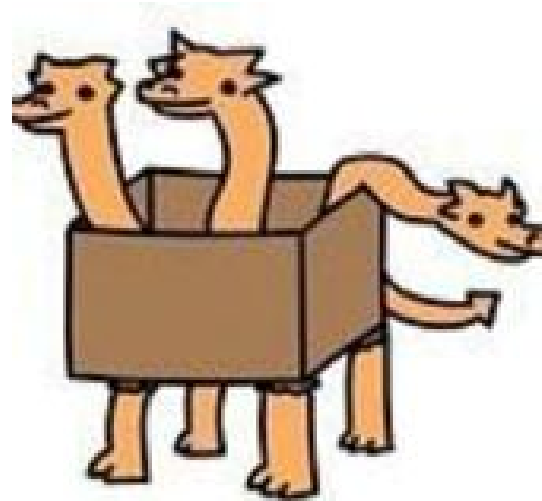
# Bayou City DAMS

- Phase 1: Systems Installation
- Phase 2: Data Migration
- Phase 3: Interface Development
- Assessment, Documentation, and Training



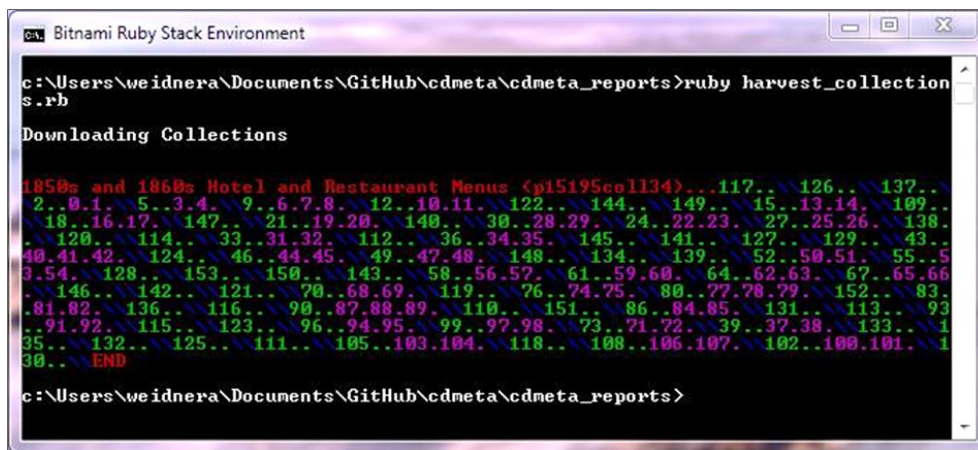
# Armand: Hydra-in-a-Box

- Community Software
- Standards Based
  - DPLA MAP
  - PCDM
- Linked Data / DPLA Hubs
- Easier Long Term Maintenance



# Buffalo: CONTENTdm Export

- Ruby Library
- CONTENTdm API
- Create Reports
- Contribute to Community



```
Bitnami Ruby Stack Environment

c:\Users\weidnera\Documents\GitHub\cdmeta\cdmeta_reports>ruby harvest_collection
s.rb

Downloading Collections

1850s and 1860s Hotel and Restaurant Menus <p15195coll34>...117...126...137...
2..0.1..5..3.4..9..6.7.8..12..10.11..122...144...149...15...13.14..109...
18..16.17..147..21..19.20..140...30..28.29..24..22.23..27..25.26..138...
120...114...33..31.32..112...36..34.35..145...141...127...129...43...
40.41.42..124...46..44.45..49..47.48..148...134...139...52..50.51..55..5
0.54..128...153...150...143...58..56.57..61..59.60..64..62.63..67..65.66
146...142...121...70..68.69..119...76..74.75..80..77.78.79..152...83...
81.82..136...116...90..87.88.89..110...151...86..84.85..131...113...93...
91.92..115...123...96..94.95..99..97.98..73..71.72...39..37.38..133...1
35...132...125...111...105..103.104..118...108..106.107..102..100.101..1
30...END

c:\Users\weidnera\Documents\GitHub\cdmeta\cdmeta_reports>
```

# Cedar: Local Vocabulary Manager

- Ruby on Rails Gem: iQvoc
- Simple Knowledge Organization System (SKOS)
- Local Customizations
  - Theming
  - ARK Minter Integration

# Cedar: Local Vocabulary Manager

Cedar

MENU

Concepts Hierarchical

Agent

Organization

Person

Place

Topic

Bayous

Bays

Beer

Bells

Benches

Benefactors

Bicycles

Bicyclists

Sections

HIERARCHICAL

ALPHABETICAL

DRAFT

EXPIRED

Links

NEW CONCEPT

# Cedar: Local Vocabulary Manager

Cedar

MENU

Bayous Concept

CREATE NEW VERSION

MAIN

LABELS

RELATIONS

NOTES

CONCEPT MAPPINGS

<http://id.loc.gov/authorities/subjects/sh85012513> *Close Matches*

<http://localhost:3001/ark:/84475/au29b56470x> *Exact Matches*

Representations

HTML

RDF/XML

RDF/TURTLE

RDF/NTRIPLES

Links

CONCEPT URI

NEW CONCEPT



# Cedar: Local Vocabulary Manager

```
@prefix : <http://localhost:3000/>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix foaf: <http://xmlns.com/foaf/spec/>.
@prefix iqvoc: <http://try.iqvoc.net/schema#>.
@prefix owl: <http://www.w3.org/2002/07/owl#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix schema: <http://localhost:3000/schema#>.
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@prefix void: <http://rdfs.org/ns/void#>.

:topic skos:prefLabel "Topic"@en.
:sh85012513 a skos:Concept;
            schema:memberOf :lcsh;
            skos:inScheme :scheme;
            skos:prefLabel "Bayous"@en;
            skos:broader :topic;
            skos:closeMatch <http://id.loc.gov/authorities/subjects/sh85012513>;
            skos:exactMatch <http://localhost:3001/ark:/84475/au29b56470x>.
```

# Greens: ID Minter & Resolver

- Ruby on Rails
- RESTful API

```
erc:  
who: Northern France  
what: Book of Hours, Use of Reims (Binding)  
when: 1400-1499  
where: http://digital.lib.uh.edu/collection/reims/item/20
```

- Mints Archival Resource Key (ARK) identifiers
- Resolves identifiers to target URLs
- Suffix Passthrough

# Greens ARK Identifier

Cedar

MENU

Bayous Concept

CREATE NEW VERSION

MAIN

LABELS

RELATIONS

NOTES

CONCEPT MAPPINGS

<http://id.loc.gov/authorities/subjects/sh85012513> *Close Matches*

<http://localhost:3001/ark:/84475/au29b56470x> *Exact Matches*

Representations

HTML

RDF/XML

RDF/TURTLE

RDF/NTRIPLES

Links

CONCEPT URI

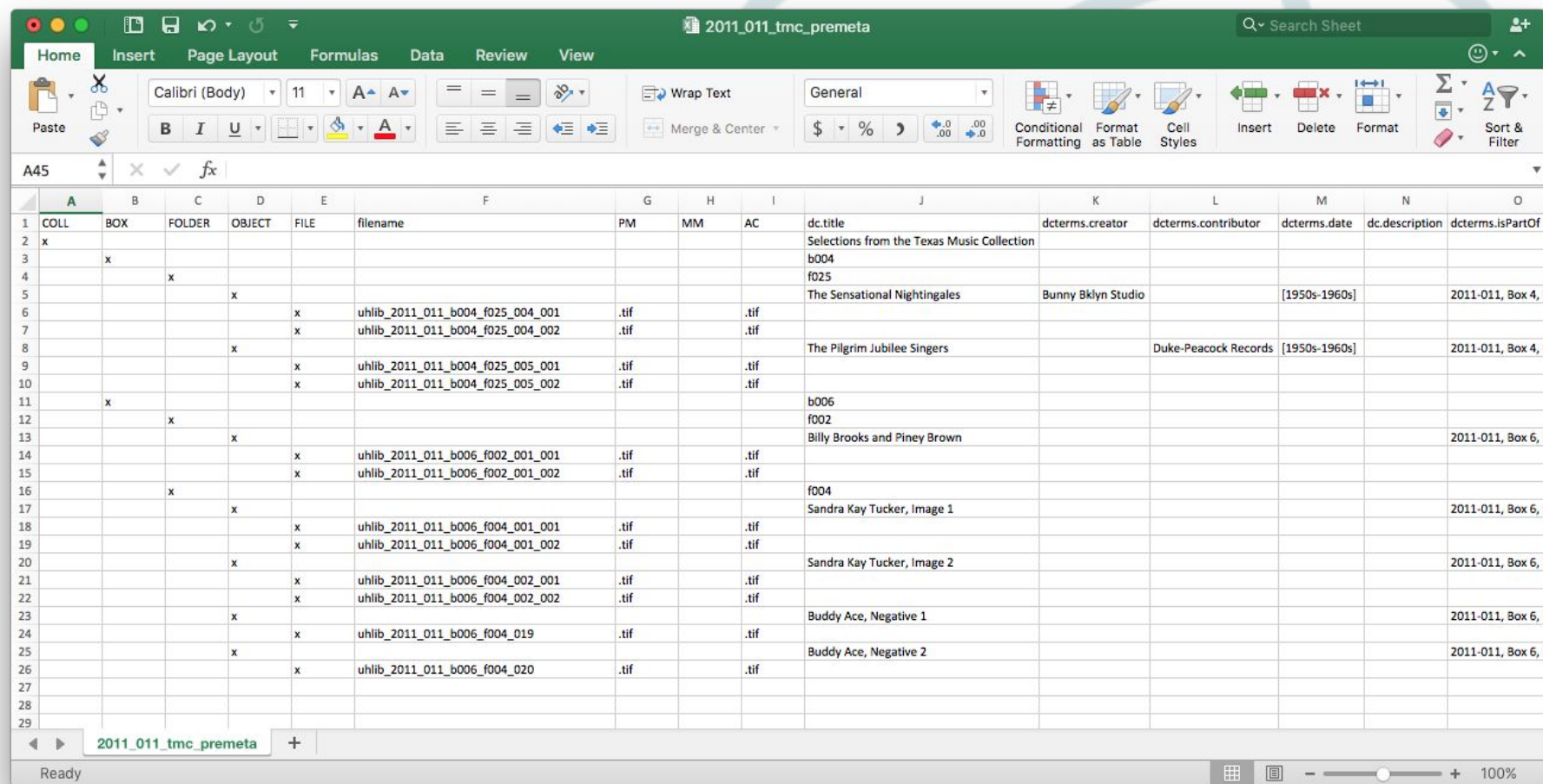
NEW CONCEPT

ark:/84475/au29b56470x

# Brays: Ingest Utility

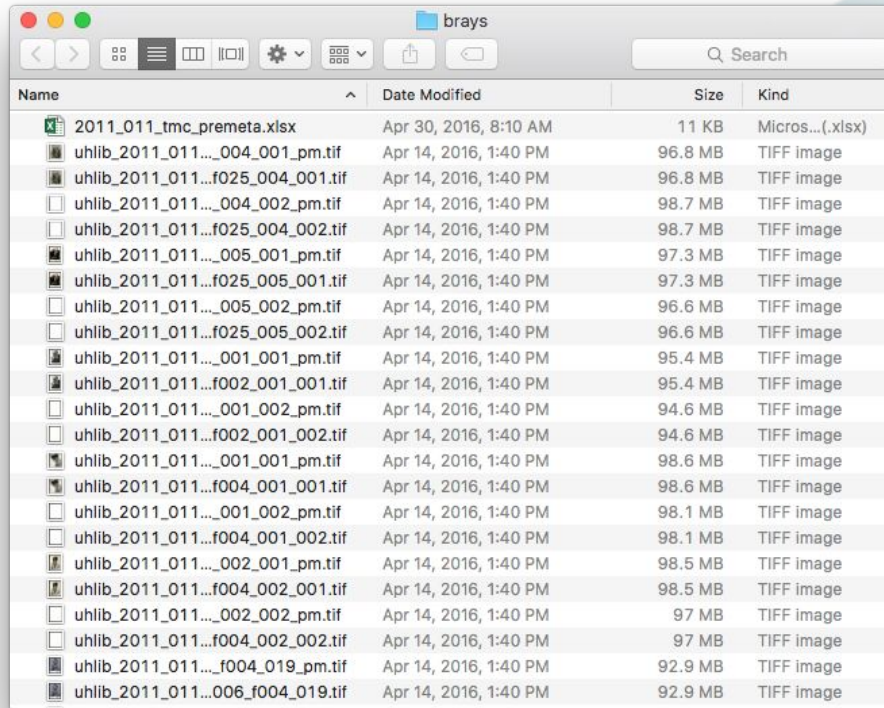
- Desktop application built off Electron framework
- Cross platform (Windows, MacOS, Linux)
- Inputs XSLT Metadata file
- Outputs Archivematica ingest package
- To-do: Hydra-in-a-box (Armand) ingest package

# Brays: Ingest Utility



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	COLL	BOX	FOLDER	OBJECT	FILE	filename	PM	MM	AC	dc.title	dcterms.creator	dcterms.contributor	dcterms.date	dc.description	dcterms.isPartOf
2	x									Selections from the Texas Music Collection					
3		x								b004					
4			x							f025					
5				x						The Sensational Nightingales	Bunny Bklyn Studio		[1950s-1960s]		2011-011, Box 4, F
6					x	uhlib_2011_011_b004_f025_004_001	.tif		.tif						
7					x	uhlib_2011_011_b004_f025_004_002	.tif		.tif						
8				x						The Pilgrim Jubilee Singers		Duke-Peacock Records	[1950s-1960s]		2011-011, Box 4, F
9					x	uhlib_2011_011_b004_f025_005_001	.tif		.tif						
10					x	uhlib_2011_011_b004_f025_005_002	.tif		.tif						
11	x									b006					
12		x								f002					
13			x							Billy Brooks and Piney Brown					2011-011, Box 6, F
14				x	x	uhlib_2011_011_b006_f002_001_001	.tif		.tif						
15					x	uhlib_2011_011_b006_f002_001_002	.tif		.tif						
16		x								f004					
17			x							Sandra Kay Tucker, Image 1					2011-011, Box 6, F
18				x	x	uhlib_2011_011_b006_f004_001_001	.tif		.tif						
19					x	uhlib_2011_011_b006_f004_001_002	.tif		.tif						
20			x							Sandra Kay Tucker, Image 2					2011-011, Box 6, F
21				x	x	uhlib_2011_011_b006_f004_002_001	.tif		.tif						
22					x	uhlib_2011_011_b006_f004_002_002	.tif		.tif						
23			x							Buddy Ace, Negative 1					2011-011, Box 6, F
24				x	x	uhlib_2011_011_b006_f004_019	.tif		.tif						
25										Buddy Ace, Negative 2					2011-011, Box 6, F
26				x	x	uhlib_2011_011_b006_f004_020	.tif		.tif						
27															
28															
29															

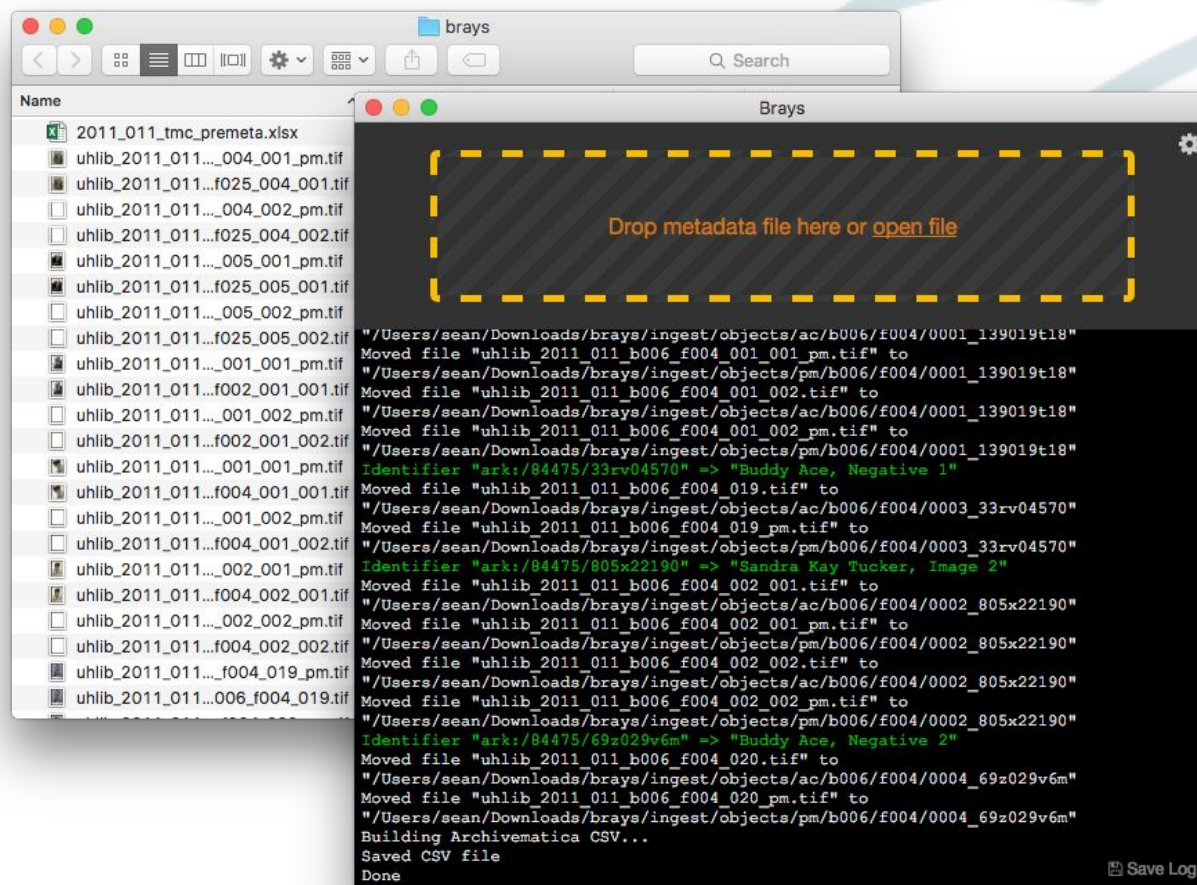
# Brays: Ingest Utility



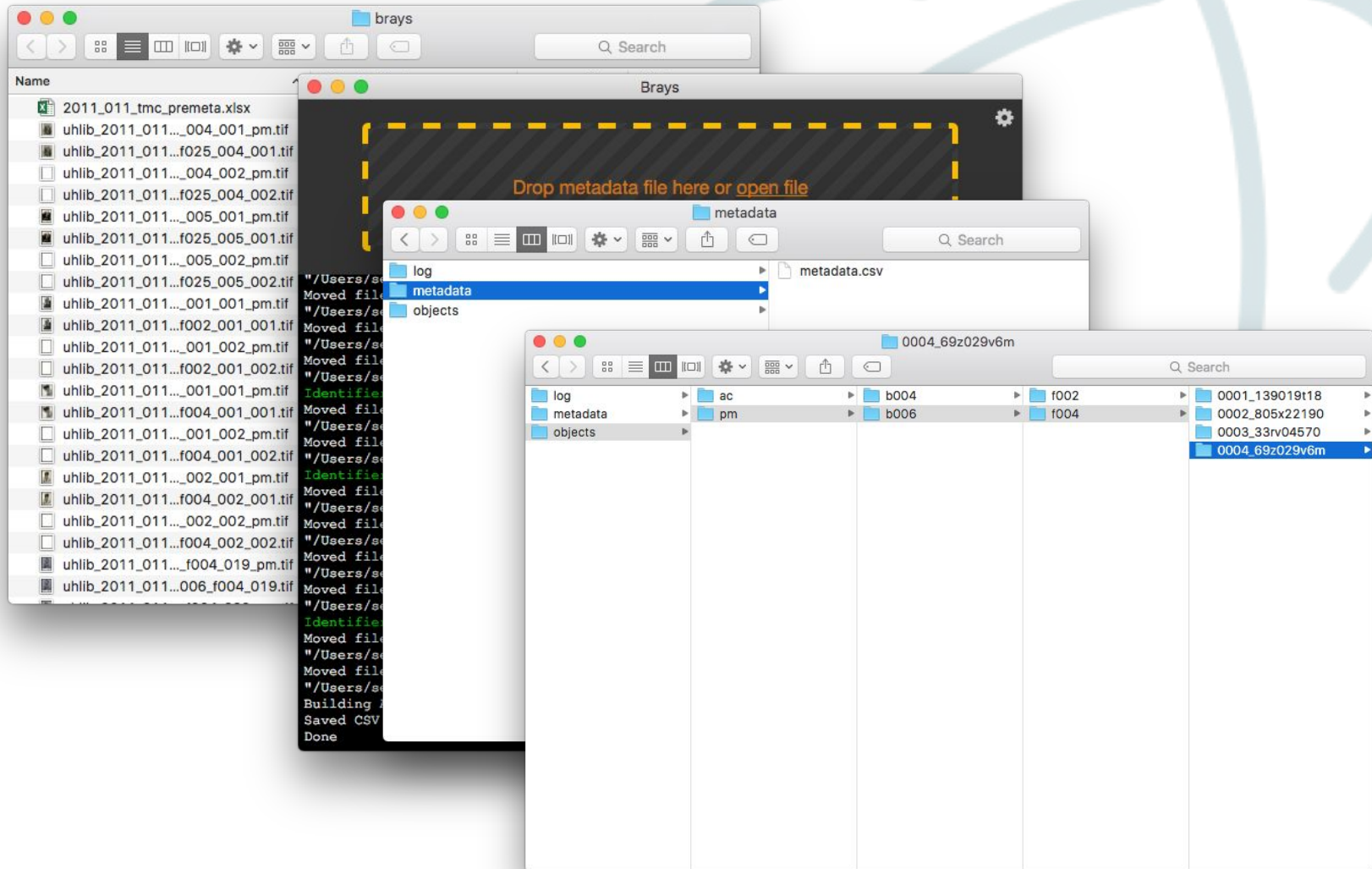
Name	Date Modified	Size	Kind
2011_011_tmc_premeta.xlsx	Apr 30, 2016, 8:10 AM	11 KB	Micros... (.xlsx)
uhlib_2011_011..._004_001_pm.tif	Apr 14, 2016, 1:40 PM	96.8 MB	TIFF image
uhlib_2011_011...f025_004_001.tif	Apr 14, 2016, 1:40 PM	96.8 MB	TIFF image
uhlib_2011_011..._004_002_pm.tif	Apr 14, 2016, 1:40 PM	98.7 MB	TIFF image
uhlib_2011_011...f025_004_002.tif	Apr 14, 2016, 1:40 PM	98.7 MB	TIFF image
uhlib_2011_011..._005_001_pm.tif	Apr 14, 2016, 1:40 PM	97.3 MB	TIFF image
uhlib_2011_011...f025_005_001.tif	Apr 14, 2016, 1:40 PM	97.3 MB	TIFF image
uhlib_2011_011..._005_002_pm.tif	Apr 14, 2016, 1:40 PM	96.6 MB	TIFF image
uhlib_2011_011...f025_005_002.tif	Apr 14, 2016, 1:40 PM	96.6 MB	TIFF image
uhlib_2011_011..._001_001_pm.tif	Apr 14, 2016, 1:40 PM	95.4 MB	TIFF image
uhlib_2011_011...f002_001_001.tif	Apr 14, 2016, 1:40 PM	95.4 MB	TIFF image
uhlib_2011_011..._001_002_pm.tif	Apr 14, 2016, 1:40 PM	94.6 MB	TIFF image
uhlib_2011_011...f002_001_002.tif	Apr 14, 2016, 1:40 PM	94.6 MB	TIFF image
uhlib_2011_011..._001_001_pm.tif	Apr 14, 2016, 1:40 PM	98.6 MB	TIFF image
uhlib_2011_011...f004_001_001.tif	Apr 14, 2016, 1:40 PM	98.6 MB	TIFF image
uhlib_2011_011..._001_002_pm.tif	Apr 14, 2016, 1:40 PM	98.1 MB	TIFF image
uhlib_2011_011...f004_001_002.tif	Apr 14, 2016, 1:40 PM	98.1 MB	TIFF image
uhlib_2011_011..._002_001_pm.tif	Apr 14, 2016, 1:40 PM	98.5 MB	TIFF image
uhlib_2011_011...f004_002_001.tif	Apr 14, 2016, 1:40 PM	98.5 MB	TIFF image
uhlib_2011_011..._002_002_pm.tif	Apr 14, 2016, 1:40 PM	97 MB	TIFF image
uhlib_2011_011...f004_002_002.tif	Apr 14, 2016, 1:40 PM	97 MB	TIFF image
uhlib_2011_011..._f004_019_pm.tif	Apr 14, 2016, 1:40 PM	92.9 MB	TIFF image
uhlib_2011_011..._006_f004_019.tif	Apr 14, 2016, 1:40 PM	92.9 MB	TIFF image



# Brays: Ingest Utility



# Brays: Ingest Utility



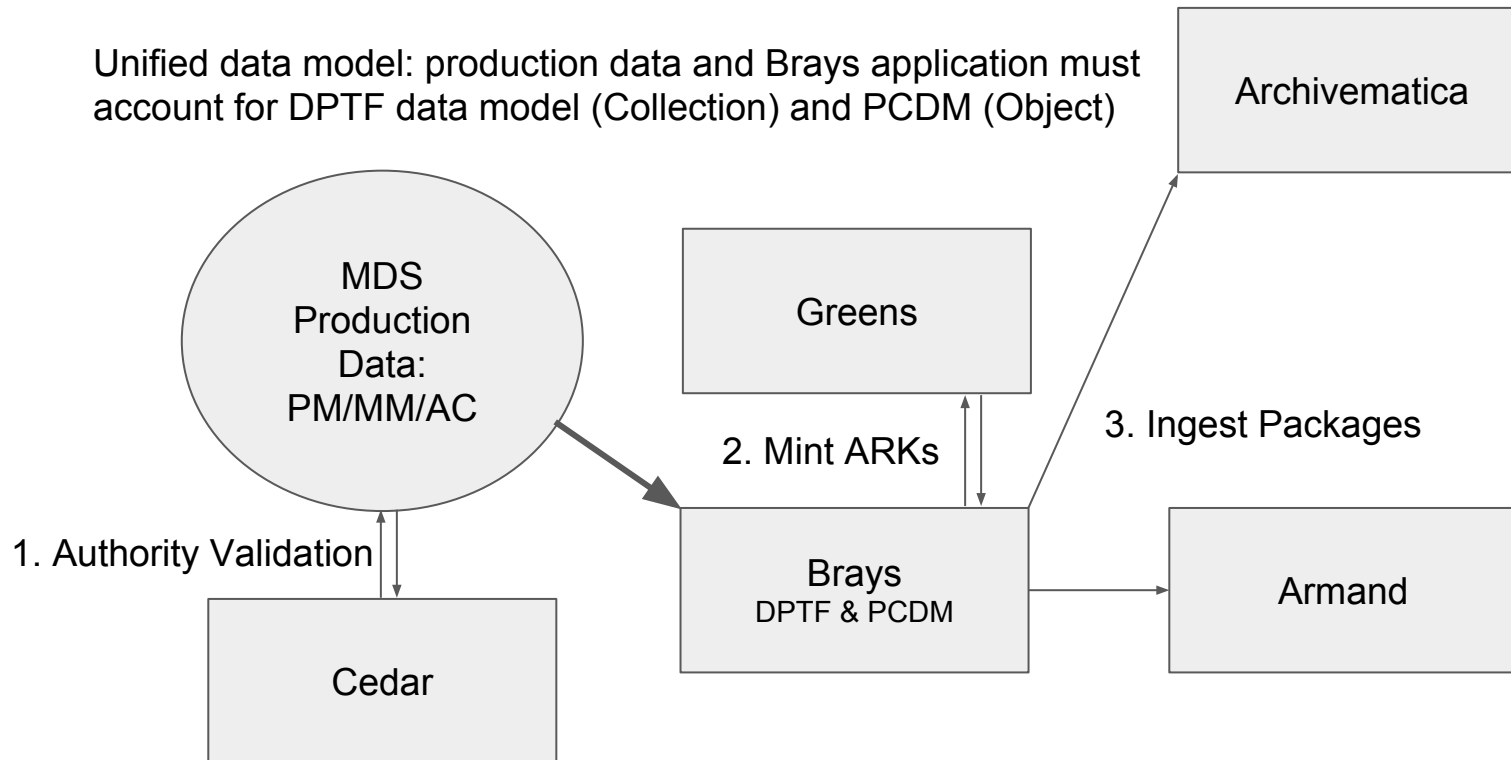
# Next Steps

*...with Sean and Bethany*

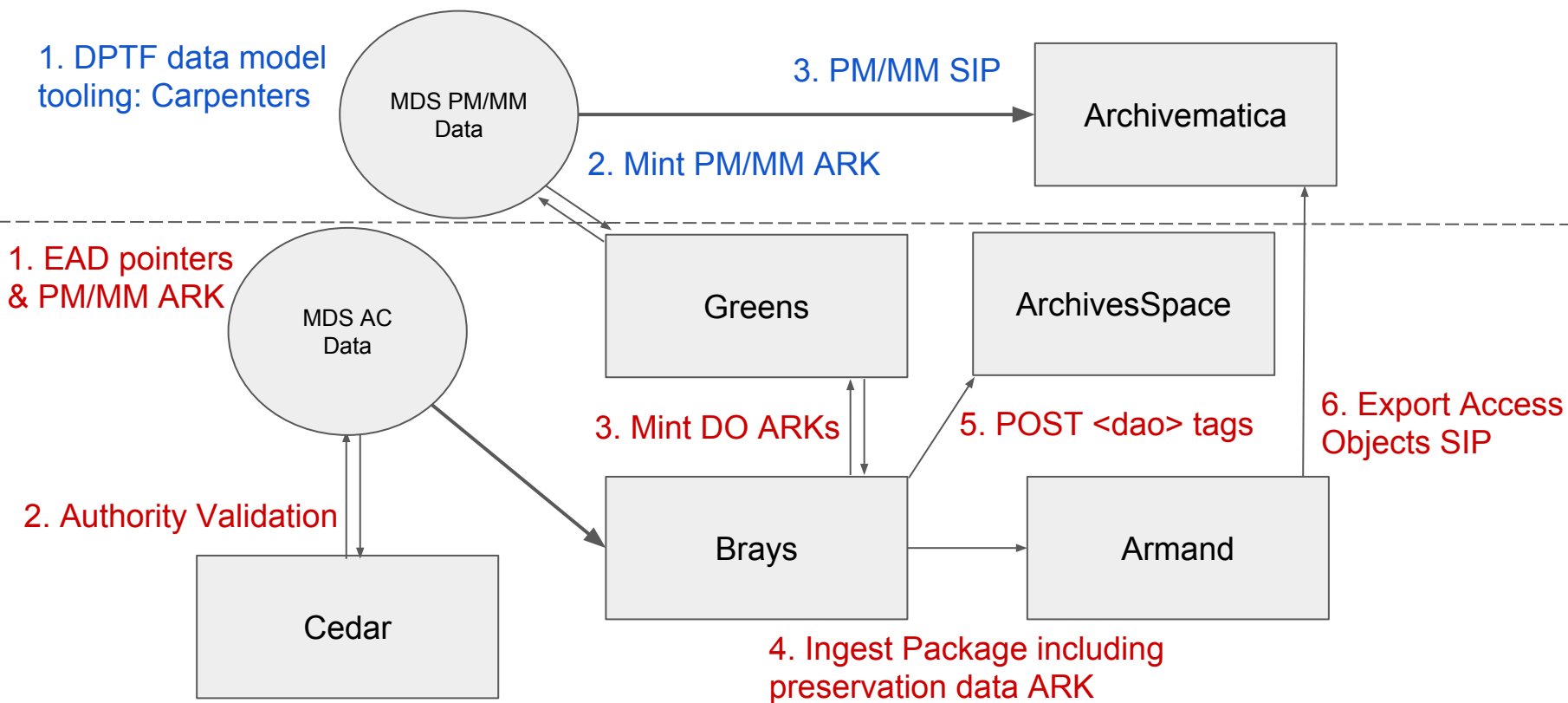
UNIVERSITY of **HOUSTON** | LIBRARIES

# Workflow Challenges

Unified data model: production data and Brays application must account for DPTF data model (Collection) and PCDM (Object)



# Next Steps: DAMS



# Next Steps: Archivematica

- Continue refining born digital workflows
- Scale up transfers with move to production instance
- Assess and update digital preservation policy



# Thank you!

Drew Krewer, Digitization Operations Coordinator, [ajkrewer@uh.edu](mailto:ajkrewer@uh.edu)

Bethany Scott, Digital Projects Coordinator, [bscott3@uh.edu](mailto:bscott3@uh.edu)

Sean Watkins, Digital Repository Developer, [slwatkins@uh.edu](mailto:slwatkins@uh.edu)

Andy Weidner, Metadata Services Coordinator, [ajweidner@uh.edu](mailto:ajweidner@uh.edu)

**GitHub:** <https://github.com/uhlibraries-digital/>