

Outside The Box: Building a Digital Asset Management Ecosystem for Preservation and Access

Current Issue

[Issue 39, 2018-02-05](#)

Previous Issues

[Issue 38, 2017-10-18](#)

[Issue 37, 2017-07-18](#)

[Issue 36, 2017-04-20](#)

[Issue 35, 2017-01-30](#)

[Older Issues](#)

For Authors

[Call for Submissions](#)

[Article Guidelines](#)

The University of Houston (UH) Libraries made an institutional commitment in late 2015 to migrate the data for its digitized cultural heritage collections to open source systems for preservation and access: Hydra-in-a-Box, Archivematica, and ArchivesSpace. This article describes the work that the UH Libraries implementation team has completed to date, including open source tools for streamlining digital curation workflows, minting and resolving identifiers, and managing SKOS vocabularies. These systems, workflows, and tools, collectively known as the Bayou City Digital Asset Management System (BCDAMS), represent a novel effort to solve common issues in the digital curation lifecycle and may serve as a model for other institutions seeking to implement flexible and comprehensive systems for digital preservation and access.

by Andrew Weidner, Sean Watkins, Bethany Scott, Drew Krewer, Anne Washington, and Matthew Richardson

Introduction

This article outlines the workflows and tools that the University of Houston (UH) Libraries have developed to facilitate digital curation activities for Hydra-in-a-Box (Hyku), Archivematica, and ArchivesSpace, collectively known as the Bayou City Digital Asset Management System (BCDAMS). BCDAMS development work began in early 2016 based on the recommendations of a UH Libraries task force charged with evaluating, testing, and implementing a new DAMS (Wu et al. 2016) [1]. The recommendation outlined a three phase implementation project, and the BCDAMS team completed most of its Phase One (Systems Installation) deliverables by the end of 2016, producing workflows and tools to support preservation and access ingest activities. As a whole, these workflows and tools comprise an ecosystem that supports the long term preservation of and access to the digitized cultural heritage materials in the UH Libraries unique collections.

Implementation Phase One

The project team, consisting of members from Digitization, Metadata, Special Collections, and Web Services, employed an agile development methodology to lay the technical foundation for the UH Libraries migration from our current systems to the BCDAMS [Figure 1].

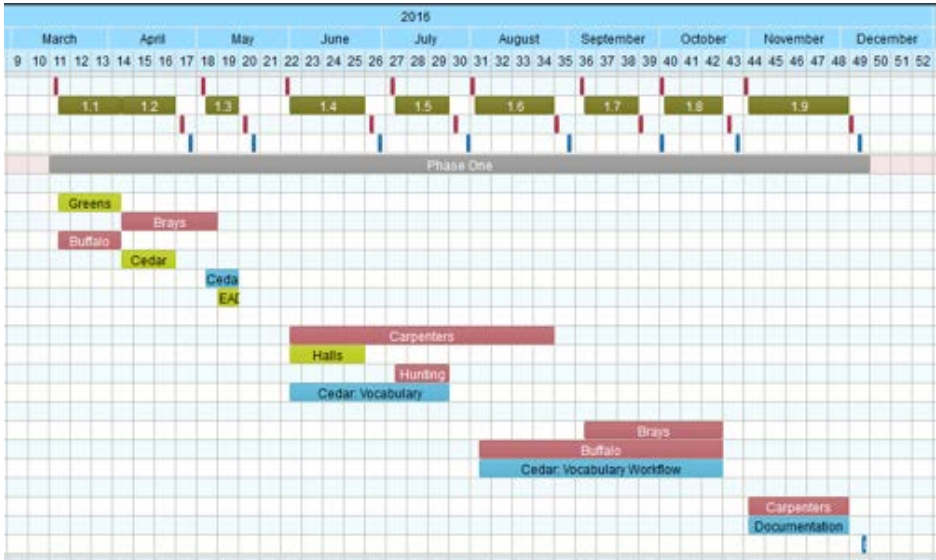


Figure 1
Bayou City DAMS Implementation Phase One. [View larger image](#)

Over nine development cycles in 2016, each lasting approximately one month, the BCDAMS team implemented five applications (named for bayous in the Houston metropolitan area) that address key digital curation tasks:

- Preparation of files and metadata for preservation ingest (**Carpenters**)
- Preparation of files and metadata for access ingest (**Brays**)
- Persistent identification of digital resources (**Greens**)
- Local management of linked data vocabularies (**Cedar**)
- Representation of archival finding aids (**Halls**)

In addition, the project team began the process of descriptive metadata analysis and remediation to prepare for collection migration, resulting in a Ruby gem for downloading metadata through the CONTENTdm API (Hunting), a framework for crosswalking metadata (Buffalo), and a metadata application profile (BCDAMS-MAP). Each development cycle ended with a formal report documenting the project’s progress that was made available to the UH Libraries, and the overall progress was recorded in a public timeline [2].

Bayou City DAMS Ecosystem

As a whole the BCDAMS applications represent an ecosystem of modular components, each described in detail in the sections below, that work together to address all aspects of the digital curation lifecycle [Figure 2]. The preservation workflow begins in Special Collections where the finding aid for a collection is created in ArchivesSpace. Using the Carpenters digitization workflow and preservation ingest application, Special Collections personnel import the ArchivesSpace finding aid and select the list of items to be digitized by checking boxes next to the folder or item in the collection hierarchy.

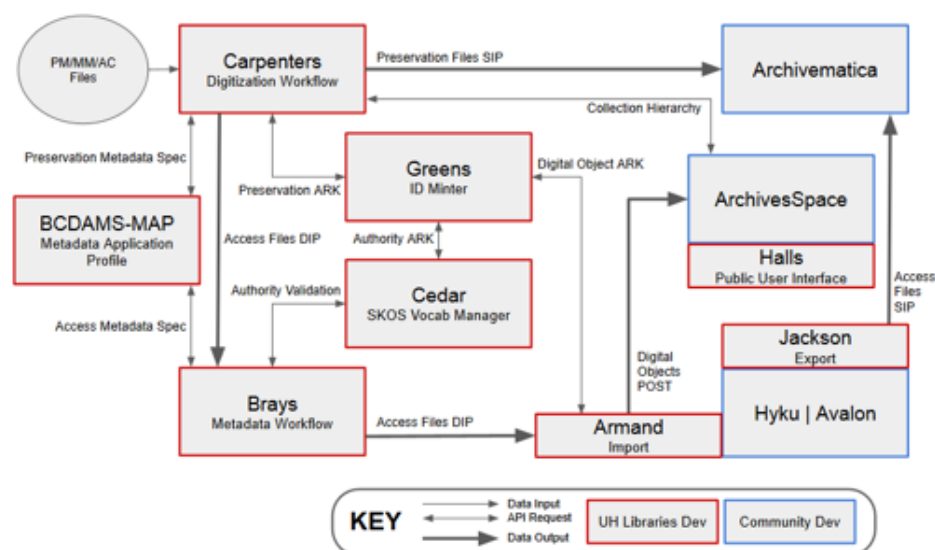


Figure 2

Bayou City DAMS Ecosystem Architecture. [View larger image](#)

As materials are scanned, Digitization Unit personnel associate files with digital objects in the collection's intellectual hierarchy within the Carpenters interface. After scanning is complete, Carpenters exports a Submission Information Package (SIP) that reflects the collection hierarchy for preservation with Archivematica. During the export, Carpenters requests a new identifier for the SIP from the Greens persistent identifier minter and produces a Dissemination Information Package (DIP) that contains the access files and minimal descriptive metadata pulled from ArchivesSpace.

The access portion of the workflow begins when Metadata Unit personnel load the Carpenters DIP in the Brays descriptive metadata editor and create descriptive metadata for all objects. Brays suggests controlled vocabulary terms from the Cedar linked data vocabulary manager and validates the record against the descriptive metadata specification defined in the BCDAMS-MAP. Future development will focus on building the tooling necessary for minting digital object identifiers, posting digital object records to ArchivesSpace, and packaging Brays output for batch ingest into the Hyku repository.

Carpenters: Digitization Workflow & Preservation Ingest

Carpenters is a cross-platform desktop application built using the [Electron framework](#). Electron uses web technologies such as HTML5, CSS, and JavaScript to create rich applications without the need to develop for specific platforms such as Windows, Mac, and Linux. Because Electron uses web technologies it can take advantage of other libraries and frameworks used in web development. As a result, the [Angular](#) and [Bootstrap](#) frameworks were used to create Carpenters' user interface and backend processes. Together, these frameworks provided quick development time, flexibility between different user groups, and easy deployment.

Carpenters allows preservation administrators to organize digitized content into hierarchies that preserve the contextual linkages and provenance of the original archival collection. This process begins in Special Collections, where curators

create a finding aid in ArchivesSpace. The provenance collection is then selected in the Carpenters Selection interface, which queries the [ArchivesSpace API](#) to automatically display the nested archival objects contained in the collection. Special Collections staff select the archival objects to be digitized using the check boxes at each hierarchy level in Carpenters, creating a shot list in the Files interface that informs the Digitization Unit which items are included in the project [Figure 3]. As materials are scanned, Digitization Unit staff associate files for each digitized item stored on the local file system with the corresponding archival object in the Files interface [Figure 4]. Through this process, the ArchivesSpace URI for each archival object is linked to its digital surrogate.

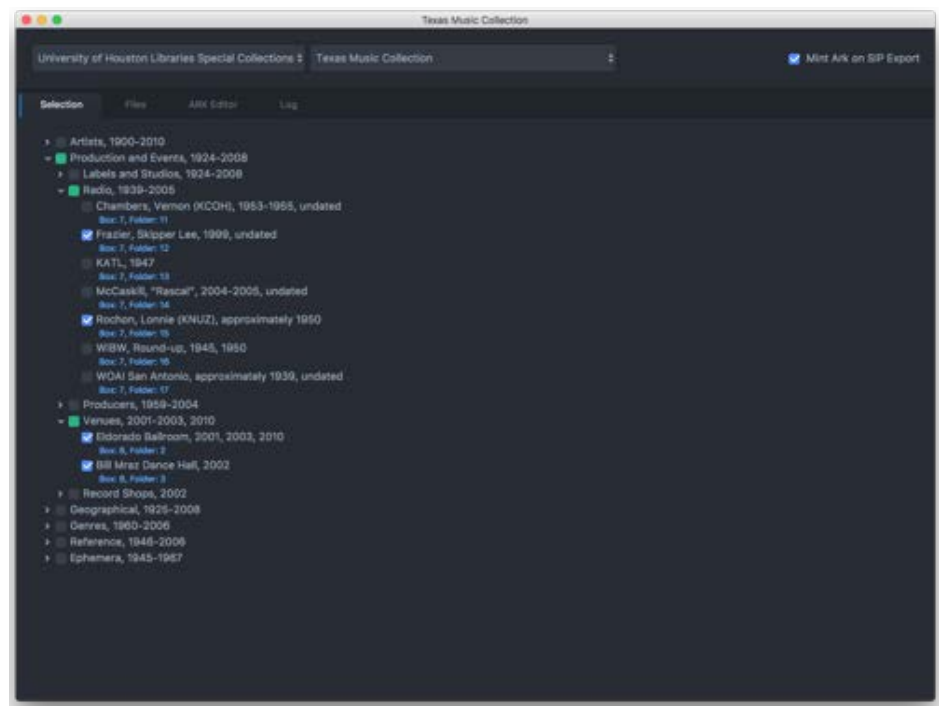


Figure 3
Carpenters selection interface with data loaded from an ArchivesSpace finding aid. [View larger image](#)

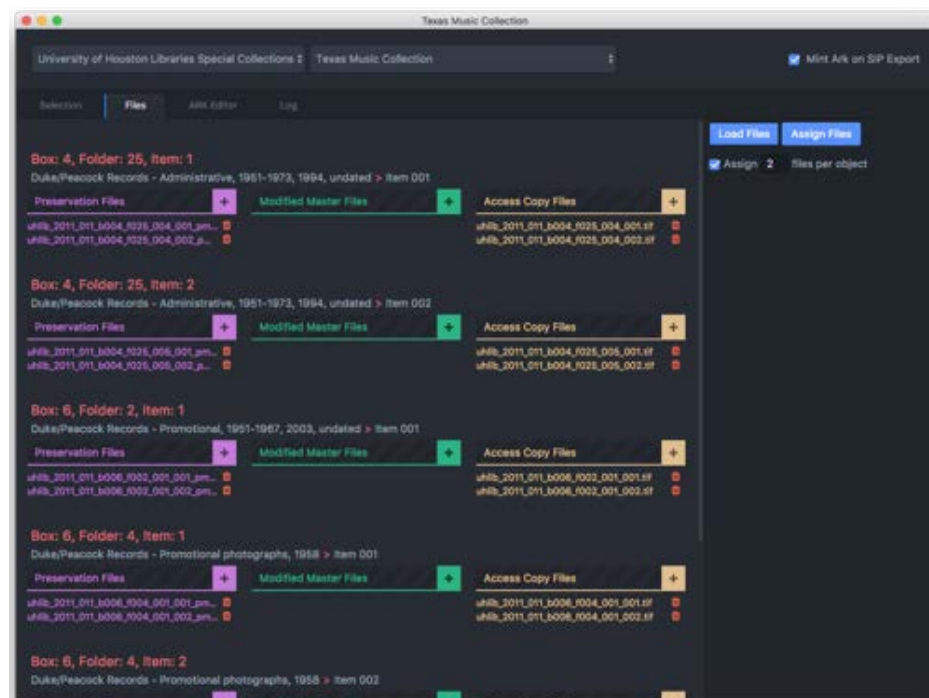


Figure 4
Carpenters file association interface. [View larger image](#)

Utilizing the arrangement provided by the ArchivesSpace-imported hierarchy structure, Carpenters automatically moves the preservation masters on the file system to a set of nested directories in an Archivematica-compatible SIP [Figure 5] that replicates the intellectual arrangement of the original collection. By eliminating the need for manually creating directories or moving files on the file system, the Carpenters application streamlines batch ingest preparation for preservation master files, making it ideal for large-scale workflows. The tool integrates with Greens to mint an Archival Resource Key (ARK) for each preservation package, creating a persistent identifier that connects the preservation master files to the access objects published in Hyku. Carpenters also outputs a DIP of access files and a metadata CSV file that is used as input for the Brays descriptive metadata application.

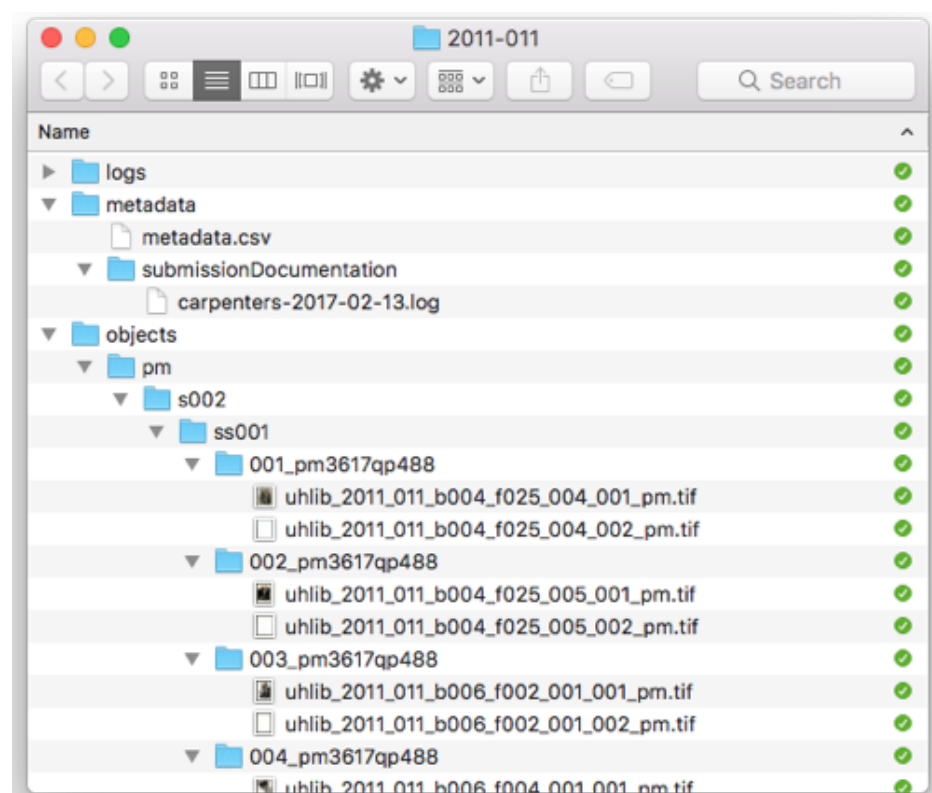


Figure 5
Archivematica SIP exported from Carpenters

Brays: Descriptive Metadata Workflow & Access Ingest

[Brays](#) provides a metadata creation interface [Figure 6] and file viewer [Figure 7] for metadata specialists working with digital objects in preparation for Hyku ingest. It is based on the same tools and frameworks found in Carpenters, and it employs a similar look and feel to give a sense of unity and connection between the two products within the BCDAMS workflow. Before Brays was developed, Metadata Unit staff used a combination of spreadsheets, Microsoft Access databases, and [AutoHotkey](#) scripts to produce tab delimited files for batch ingest. The BCDAMS implementation team held focus groups with metadata specialists to gather information about their workflow and solicit feature requests for a new descriptive metadata creation tool. The result is an application that promises to greatly improve efficiency in the descriptive metadata creation workflow through a one-stop interface that integrates with other BCDAMS applications.

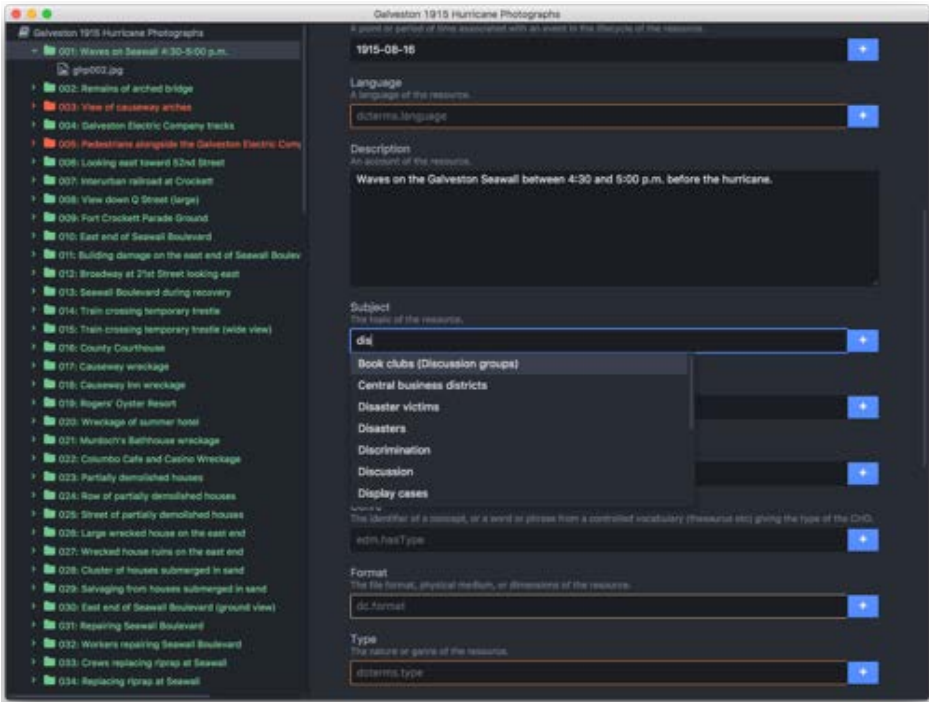


Figure 6
Brays metadata creation interface with validation and type-ahead suggestions.
[View larger image](#)

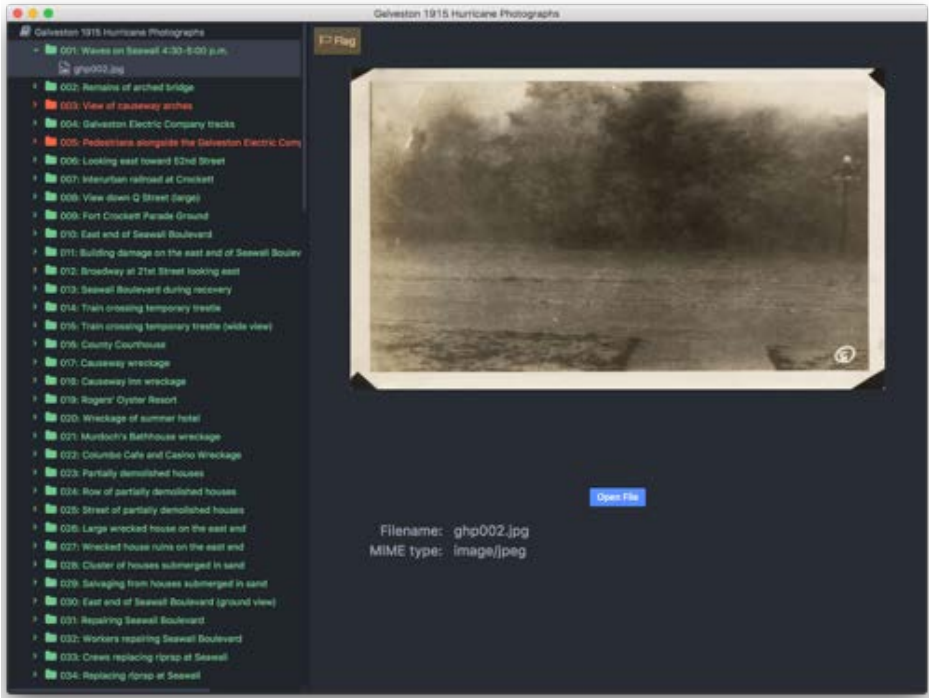


Figure 7
Brays file viewer. [View larger image](#)

Brays dynamically reads and writes to a metadata CSV file included in the DIP generated by the Carpenters application described above. Color coding in the metadata creation interface indicates to metadata specialists which fields are

required, recommended, and optional [3]. Additionally, once the record contains all required fields, the object name in the object viewer turns from red to green. Problematic records can be flagged and annotated to record contextual information during the metadata creation process. Metadata specialists can also use the autofill function to populate values that are repeated in all or a select group of records. Entries in the date field are validated against ISO/WD 8601-2 [4], the proposed Extended Date/Time Format (EDTF) extension of ISO 8601. Inspired by the [EDTF Humanize gem](#), Brays parses [EDTF dates](#) and translates them into a human readable form in the metadata entry display.

In order to ensure metadata consistency across systems, Brays retrieves the metadata elements in the [BCDAMS-MAP](#) with a query to the access metadata specification API [5]. Brays applies those requirements to enforce valid metadata entry. For some metadata fields, Brays leverages ranges defined in the MAP to provide metadata specialists with a controlled list of authorized field values. Other fields require values from the Cedar vocabulary manager (see below), which are presented as suggestions when metadata specialists type in the form field.

Future plans for Brays development include an export mechanism for packaging the access and metadata files into a format suitable for ingest into our Hyku repository.

Greens: Persistent Identifier Minter & Resolver

Integral to the Brays and Carpenters workflow and ingest applications are tools that constitute the BCDAMS ecosystem infrastructure: the Greens persistent identifier minter and the Cedar linked data controlled vocabulary manager (described in the next section). The BCDAMS team recognized early on that the persistent identification of resources is essential to maintaining the long-term integrity of digital collections. As systems change and data is migrated, the BCDAMS must be able to provide seamless access to the resources it manages. The team selected the ARK specification (Kunze and Rogers 2008) [6] as the persistent identifier scheme most applicable to digital resources for cultural heritage collections. After investigating the [EZID service](#), the team chose to implement its own identifier minter and resolver to manage identifiers locally.

[Greens](#) is a Ruby on Rails application that mints opaque ARK identifiers with the [NOID gem](#) and includes the ability to add prefixes. The BCDAMS team plans to mint identifiers for three types of resources with the following prefixes:

- pm: preservation master SIPs
- do: digital objects
- au: authority records [Figure 8]

```
https://id.lib.uh.edu/ark:/84475/au9180vs53d
erc:
who: University of Houston Libraries Vocabularies
what: Bayous
when: 2016-06-23
where: https://vocab.lib.uh.edu/en/concepts/sh85012513
```

Figure 8

ARK identifier and Greens ERC record for a BCDAMS authority.

Greens resolves ARKs, passing through any suffixes appended to the identifier, and the API supports minting, updating, and destroying ARKs managed by the application. Two BCDAMS applications currently depend on Greens to mint and update ARKs: Carpenters (preservation package identifiers) and Cedar (authority identifiers).

Cedar: Linked Data Vocabulary Manager

Cedar is a Ruby on Rails application driven by the [iQvoc SKOS vocabulary management gem](#). The BCDAMS team has started to populate a local Cedar instance with vocabulary terms already in use across UH Libraries systems: the [UH Digital Library](#), the [UH Institutional Repository](#), and the [finding aids in Special Collections](#). The UH Libraries local SKOS vocabulary is based on the context classes of the [DPLA MAP](#). As mentioned above, Brays makes use of Cedar data [7] through the [iQvoc API](#) for authority control during the descriptive metadata creation process.

Records in Cedar are served as HTML for end users and RDF/XML, N-Triples, and Turtle [Figure 9] machine-readable formats. Whenever possible, UH Libraries authorities link to external vocabularies such as the Library of Congress Subject Headings and the Getty Thesaurus of Geographic Names. Each record linked to an external authority is organized in a SKOS Collection for that vocabulary, and every external authority link uses the SKOS `closeMatch` predicate. The SKOS `exactMatch` predicate is reserved for the authority ARK identifier minted in Greens.

```
@prefix : <https://vocab.lib.uh.edu/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix schema: <https://vocab.lib.uh.edu/schema#>.
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.

:concept skos:prefLabel "Concept"@en.
:sh85012513 a skos:Concept;
  schema:memberOf :lcsh;
  schema:memberOf :aat;
  skos:inScheme :scheme;
  skos:prefLabel "Bayous"@en;
  skos:broader :concept;
  skos:closeMatch <http://id.loc.gov/authorities/subjects/sh85012513>;
  skos:closeMatch <http://vocab.getty.edu/aat/300008674>;
  skos:exactMatch <https://id.lib.uh.edu/ark:/84475/au9180vs53d>.
```

Figure 9

Turtle representation of a Cedar authority record. [View larger image](#)

Halls: Finding Aid Representation

In an effort to envision complementary access systems for digitized archival objects, the BCDAMS team and Special Collections recognized the benefits of developing a custom public interface for finding aids managed in [ArchivesSpace](#). Early experiments using XSLT to transform EAD for display on the web [8] helped to gain buy-in for this approach. The decision to pursue local development of an

ArchivesSpace public interface was ultimately guided by a desire to ensure that we implement a finding aid access system driven by users' research needs [9]. This path also presents an opportunity to maximize systems integration and provide a seamless user experience between the BCDAMS access repository and the finding aid interface.

By displaying the context of the archival or provenance collection for content stored in the digital repository [Figure 10], Halls will play a key role in the access portions of the larger BCDAMS architecture. Currently a proof-of-concept slated for further development, Halls uses XSLT transformations of EAD-encoded finding aids pulled from the ArchivesSpace API to present digital objects in the Hyku repository within their archival context. Halls will also display search results that clearly convey both the intellectual arrangement (series, sub-series, and file-level titles and descriptions) and the physical instance (box and folder information) associated with an archival object.

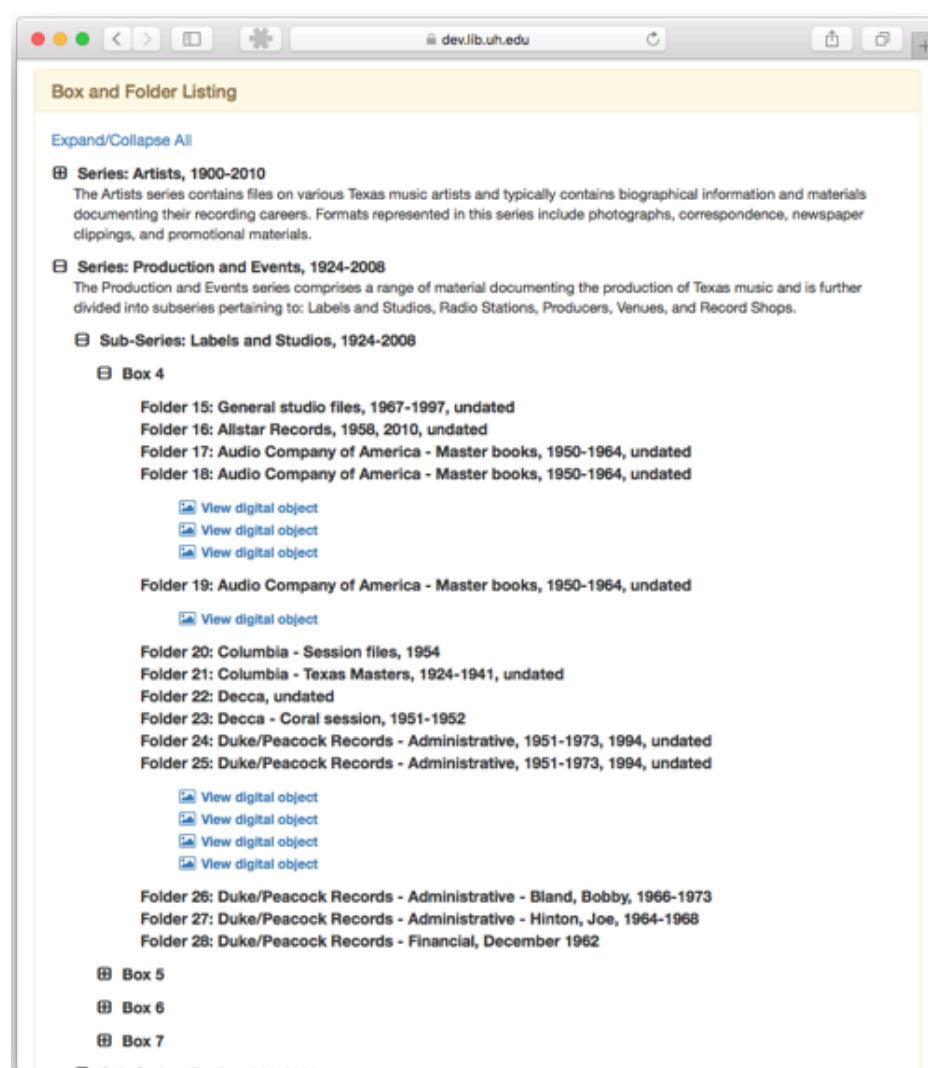


Figure 10

Halls finding aid hierarchy with links to the digital object. [View larger image](#)

Beyond the Halls finding aid interface, ArchivesSpace and its API allow for other integrations within the BCDAMS digital curation ecosystem. The archival arrangement managed in ArchivesSpace informs the hierarchies that are used to

structure a digital collection's preservation SIP in Carpenters, and archival description provides a starting point for metadata creation in Brays. With the goal of leveraging existing archival arrangement and description in the overall workflow for both preservation and access, we are approaching the migration of finding aid data from Archon to ArchivesSpace as a data quality enhancement and clean-up project. Many legacy finding aids in Archon predate standards such as [Describing Archives: a Content Standard](#) or [Encoded Archival Description](#). As each finding aid is imported into ArchivesSpace, a careful review is necessary to ensure quality and consistency. While this was less of an issue in the previous, siloed ecosystem, standardized arrangement and description—including adherence to clearer input guidelines—is crucial for sharing and reusing archival metadata across the BCDAMS digital curation ecosystem.

Next Steps

Phase Two of BCDAMS implementation calls for the migration of digital objects from the UH Digital Library to the BCDAMS. As with our collection finding aids, migrating digital objects to a new system presents opportunities for improving metadata quality at many levels. Toward that end, the BCDAMS team developed the [Hunting Ruby gem](#), which provides convenience methods to access descriptive metadata through the [CONTENTdm API](#). Combined with a custom framework for crosswalking metadata called [Buffalo](#), Hunting allows metadata creators to quickly preview how the mapped data will appear in the new system and produce reports that facilitate data cleanup and controlled vocabulary analysis.

The BCDAMS-MAP is a central component of the migration effort that is already underway. Two ad hoc working groups, the [Descriptive Metadata Working Group](#) and the [Rights Metadata Working Group](#), are using the Github wiki and issue tracker to develop the BCDAMS-MAP metadata element set and crosswalk. The BCDAMS-MAP, a website built with the [Jekyll static site generator](#), doubles as a host for the tables used to preview crosswalked metadata [10]. Just as the [UH Libraries Metadata Dictionary](#) served to guide metadata creation for the UH Digital Library, the [BCDAMS-MAP input guidelines](#) will assist metadata specialists as they modify existing legacy data for migration to the BCDAMS.

After the initial release of [Hyku](#), the BCDAMS team will engage in thorough ingest and migration testing of the system with the goal of deploying a fully functional production workflow and demonstration repository that can be used to solicit stakeholder feedback. Development activities will continue with the creation of the Armand (ingest) and Jackson (export) utilities that will take advantage of the API hooks being developed in Hyku. Upon importing a DIP to the repository, Armand will mint identifiers for each object and POST a corresponding digital object to ArchivesSpace at the appropriate level of description. End users will then be able to view digital objects in their archival context through the Halls finding aid interface. Jackson will create a preservation SIP containing the Hyku access objects for Archivematica ingest. The BCDAMS team will also begin investigating and experimenting with the [Avalon Media System](#) as an access solution for audio-visual materials.

Conclusion

With this article, the BCDAMS team hopes to inform DAMS migration work underway at many institutions and provide a model for institutions considering

migration projects. The tools that the BCDAMS team has developed address some of the many challenges associated with access, preservation, persistent identification, authority control, and archival representation in the cultural heritage digitization context. By sharing these tools openly and widely, we hope to create opportunities for conversation and collaboration around these common challenges.

References

- [1] Wu A, Thompson S, Vacek R, Watkins S, Weidner A. 2016. Hitting the Road towards a Greater Digital Destination: Evaluating and Testing DAMS at the University of Houston Libraries. *Information Technology and Libraries*. [Cited 2017 Feb 15]; 35(2). Available from: <http://ejournals.bc.edu/ojs/index.php/ital/article/view/9152>
- [2] Bayou City DAMS Project Timeline. [Cited 2017 Feb 15]. Available from: <https://www.tomsplanner.com/public/bayou-city-dams>
- [3] Future development will include additional visualizations beyond color coding.
- [4] Data elements and interchange formats — Information interchange — Representation of dates and times — Part 2: Extensions (ISO/WD 8601-2). International Organization for Standardization. [Cited 2017 Mar 10]. Available from: http://www.loc.gov/standards/datetime/iso-tc154-wg5_n0039_iso_wd_8601-2_2016-02-16.pdf
- [5] BCDAMS-MAP Brays API. University of Houston Libraries. [Cited 2017 Feb 15]. Available from: <https://vocab.lib.uh.edu/bcdams-map/api/brays.json>
- [6] Kunze J, Rogers R. 2008. The ARK Identifier Scheme. California Digital Library. [Cited 2017 Feb 15]. Available from: <https://escholarship.org/uc/item/9p9863nc>
- [7] University of Houston Libraries Vocabularies. [Cited 2017 Feb 15]. Available from: <https://vocab.lib.uh.edu>
- [8] EAD-XSLT. Sean Watkins. [Cited 2017 Mar 10]. Available from: <https://github.com/seanlw/ead-xslt>
- [9] Pshock D. 2017. Finding aids user interviews, December 2016. User Experience at University of Houston Libraries. [Cited 2017 Feb 15]. Available from: <http://weblogs.lib.uh.edu/ux/finding-aids-interviews>
- [10] BCDAMS-MAP Migration. University of Houston Libraries. [Cited 2017 Feb 15]. Available from: <https://vocab.lib.uh.edu/bcdams-map/migration>

About the Authors

Andrew Weidner is the Metadata Services Coordinator at the University of Houston (UH) Libraries where he oversees the Metadata Unit and is the project manager for the Bayou City DAMS. Prior to joining UH, Andrew worked as the New Mexico state project coordinator for the National Digital Newspaper Program at the University of North Texas (UNT). He holds masters degrees in Library Science from UNT and History from the University of Texas of the Permian Basin. Email: ajweidner@uh.edu

Sean Watkins is the Lead Repository Developer at University of Houston Libraries. Sean earned his degree in Computer Science at Southeast Missouri State University in 2002. Sean has developed and lead many web application projects, which include system integrations and data mapping to provide better experiences for users. When Sean isn't glued to a computer screen he spends his time playing board games, video games and taking on small home automation projects. He currently serves on the DAMS Task Force and is responsible for the system architecture and development of the new digital library.
Email: slwatkins@uh.edu

Bethany Scott is the Coordinator of Digital Projects at the University of Houston Libraries where she currently works on projects to implement systems such as ArchivesSpace, Archivematica, BitCurator, and Omeka. As a representative of UH Special Collections, she contributes knowledge on digital preservation, born-digital archives, and archival description to the DAMS Task Force. She received an MS in Information Studies from the University of Texas at Austin and a BA in Studio Art from the University of St. Thomas.
Email: bscott3@uh.edu

Drew Krewer is the Digitization Services Coordinator for the University of Houston Libraries, where he also serves as chair of the Digital Preservation Task Force. He is currently the chair of the DPN Preservation Metadata Standards Committee. He holds both an MFA in Poetry and an MA in Information Resources and Library Science from the University of Arizona and has worked in the past with the Getty Research Institute, the Center for Creative Photography, and the University of Arizona Poetry Center.
Email: ajkrewer@uh.edu

Anne Washington is the Metadata Librarian at the University of Houston Libraries where she manages the Metadata Unit, which is responsible for creating and improving digital collections metadata. Her research interests include linked data initiatives with a focus on description and discovery strategies. She is currently the co-chair of the ALCTS/LITA Linked Library Data Interest Group. Anne earned her MLIS from the University of Wisconsin-Milwaukee and held previous positions at the University of Virginia Library and the University of Wisconsin-Milwaukee Libraries.
Email: awashington@uh.edu

Matthew Richardson is Program Manager in University of Houston Libraries Special Collections. There he is responsible for managing archival finding aids, coordinating both physical and born digital accessions, and assisting with reference in the reading room. Current finding aid projects include the transition to ArchivesSpace, updating holdings in Texas Archival Resources Online (TARO), and coordinating description with the BCDAMS project. He holds an MA in Liberal Studies from Northwestern University and is currently pursuing an MS-LS from the University of North Texas.
Email: rmrichardson@uh.edu

Subscribe to comments: [For this article](#) | [For all articles](#)

Leave a Reply


Name (required)

Mail (will not be published) (required)

Website

☐

I'm not a robot


reCAPTCHA
[Privacy](#) - [Terms](#)

Log in

This work is licensed under a [Creative Commons Attribution 3.0 United States License](#).

