

# DEVELOPING THE DPLA TEXAS SERVICE HUB, V.2.

---

A collaboration to build the next generation aggregator for Texas

*Kristi L. Park*, Executive Director, Texas Digital Library

*Ryan Steans*, Assistant Director, Texas Digital Library

*Dr. Martin Halbert*, Dean, University of North Texas Libraries

*Judith Hiott*, Chief, Houston Area Library Network

*Nick Woodward*, Software Engineer, Texas Digital Library

*Hannah Tarver*, Supervisor of Digital Projects, University of North Texas Libraries



## A Wealth of Knowledge

explore 16,275,285 items from libraries, archives, and museums



## Exhibitions

[View all »](#)



Explore  
by Place

[Map »](#)

Explore  
by Date

[Timeline »](#)



1946 1947 1948 **1949** 1950 1951 1952



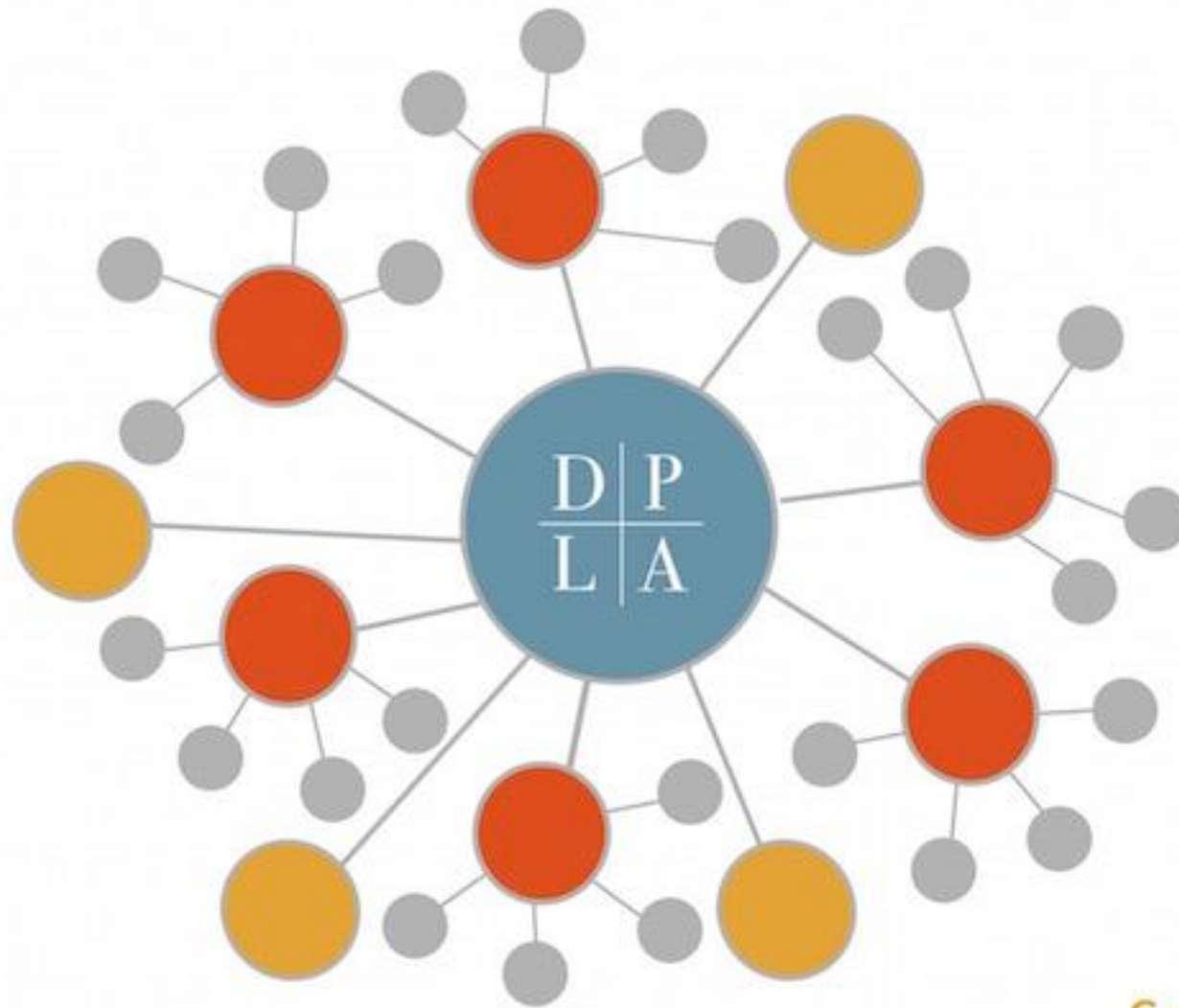
DIGITAL PUBLIC LIBRARY  
OF AMERICA

- Search & Browse portal
- Open API
- Primary Source Sets for K-12 and Higher Ed
- Hydra in a Box (Hyku)
- Ebooks initiatives

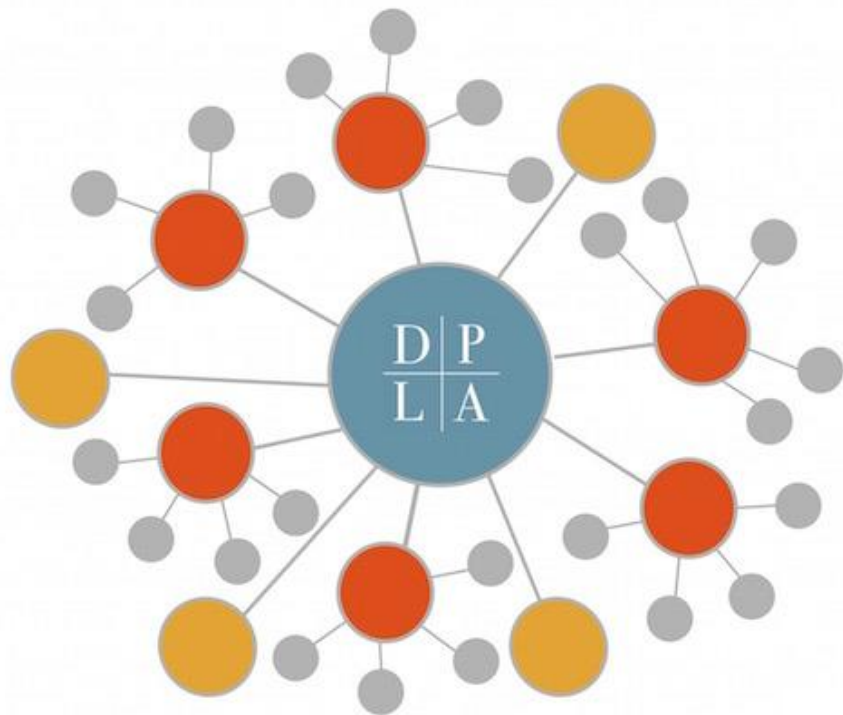


DIGITAL PUBLIC LIBRARY  
OF AMERICA

- Increase Visibility of Resources
- Promotes Information Literacy
- Enables free dissemination of information
- Connects disparate collections



Content Hubs  
Service Hubs



**MWDL**  
Mountain West Digital Library

<http://mwdl.org/>

**MDL** MINNESOTA  
DIGITAL LIBRARY

<https://mndigital.org/>



<https://padigital.org/>

# The Portal to Texas History

Search 826,816 items from 346 partners across Texas.

Any Type Search

Sample Searches: "Breckenridge American" "Houston Post" "Abilene Daily Reporter" "Attorney General Opinion" "B-24 Liberator"

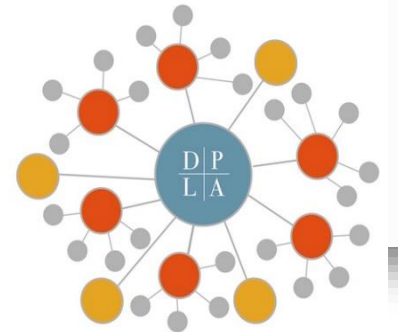
A gateway to rare, historical, and primary source materials from or about Texas:

About the Portal

Search

Explore by

Take a Tour



<https://texashistory.unt.edu/>



Damon, Katy & Damon, Fred. [Barbara Jordan Being Interviewed], photograph, October 21, 1971; ([texashistory.unt.edu/ark:/67531/metapth389285/](https://texashistory.unt.edu/ark:/67531/metapth389285/); accessed April 18, 2017), University of North Texas Libraries, The Portal to Texas History, [texashistory.unt.edu](https://texashistory.unt.edu/); crediting Texas Southern University.



DIGITAL PUBLIC LIBRARY  
OF AMERICA

Home Exhibitions Map Timeline Apps

« back to search results

Search the Library

## [Barbara Jordan Being Interviewed]



Get full image from The  
Portal to Texas History

Creator	Damon, Katy; Damon, Fred
Created Date	1971-10-21
Partner	The Portal to Texas History
Contributing Institution	The Portal to Texas History
Description	Photograph of Barbara Jordan being interviewed by a reporter at the Rice Hotel in Houston, Texas. 1 photograph : b 10 x 8 in.
Location	United States - Texas - Harris County - Houston

**UNT**<sup>®</sup>  
UNIVERSITY  
OF NORTH TEXAS

So what's the problem?

# HARRY RANSOM CENTER

THE UNIVERSITY OF TEXAS AT AUSTIN

[Collections](#) :: [Research](#) :: [Exhibitions](#) :: [Events](#) :: [Visit](#) :: [About](#)

[Home](#) [Browse All](#) [Materials Use Policy](#) [Digital Repro](#)

Search

[Home](#) > [Gabriel García Márquez Collection](#)

## Gabriel García Márquez Collection



Houston  
Preser

[Home](#) [Browse All](#) [Exhibits](#)

Search

Add or remove other collections to your search:

- ☒ Architectural Drawings
- ☒ Archival Collections
- ☒ Astrodome Memories
- ☒ Barnstone Project
- ☒ Books, Reports and Periodicals

[Show 37 more](#)

Narrow your search by:

- [Creator](#)
- [Date](#)
- [Subject](#)
- [Type](#)



You've searched: All Collections

All fields: MSS0337

Sort by: [Relevance](#)

Display: 20 ☐ Ch



☐ Louis Armstrong

### Louis Armstrong

Collection: Image Collections



#### Subject Terms:

Brown, Jewel ;  
Armstrong, Louis, 1901-1971 ;  
African American jazz musicians ;  
Jazz singers ;  
Jazz musicians

#### Date Original:

#### Description:

Louis Armstrong standing at a microphone during a performance.



TEXAS STATE LIBRARY  
AND  
ARCHIVES COMMISSION

## TEXAS DIGITAL ARCHIVE

[Home](#) > Welcome to the Texas Digital Archive

Search Archive

The Texas Digital Archive (TDA) manages, preserves, and facilitates access to the electronic records collections of the Texas State Library and Archives Commission, including those transferred by State agencies or digitized by the State Archives. All records visible in this portal are unrestricted and available for public use.

[Click here for tips on Locating records in the TDA](#)

## Collections



Governor's  
Records –  
Rick Perry,  
2000-2014



Manuscripts  
Collections



Prints and  
Photographs  
Collections



Texas  
Legislature



[Access State](#)



[Sam](#)



[Other](#)



Ship Name: Star of Peru; Sailed: 1863-1926 but still afloat in 1948; Type: Iron 3-masted later bark; Built by: Sunderland, England by Pile & Hay for J. D. Tyser & Co.; Dimensions: 201.6' x 33' x 20.5'; Tonnage: 1027 tons.

URI

<http://hdl.handle.net/2346/47417>

Collections

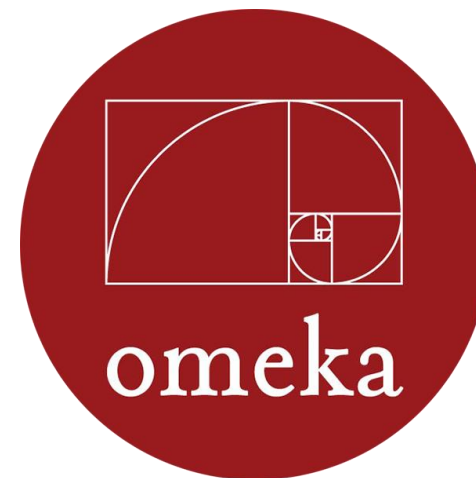
[Dyal Ships](#)

View/Open

- [Star of Peru, Himalaya, Bougainville-At sea. Star of Peru, Himalaya, Bougainville. Bark Star of Peru at sea. \(1.980Mb\)](#)
- [Star of Peru: Narrative \(24.53Kb\)](#)

Date  
2008

Author  
Dyal, Donald H.







---

LSTA Cooperation Grant funding through:



TEXAS STATE LIBRARY  
AND  
ARCHIVES COMMISSION

# Project objectives

1. Develop comprehensive inventory of major digital (archival) collections in Texas.
2. Develop a sustainable model for operations and governance for a statewide DPLA aggregation hub.
3. Implement a demonstration hub that aggregates collections from the Project Partners' digital collections, as well as a select group of other participants.

<https://tdl.org/2016/08/texas-digital-library-awarded-library-cooperation-grant/>

# Research

- Digital Collections Survey & Inventory

Q7 Does your institution currently maintain digital collections?

- ☐ Yes (1)
- ☐ No (2)
- ☐ I don't know/Not sure (3)

Q8 If you don't currently maintain digital collections, do you plan to create digital collections?

- ☐ Yes (1)
- ☐ No (2)
- ☐ I don't know/Not sure (3)

Q9 Select from the list below the types of digital collections your institution maintains (select all that apply).

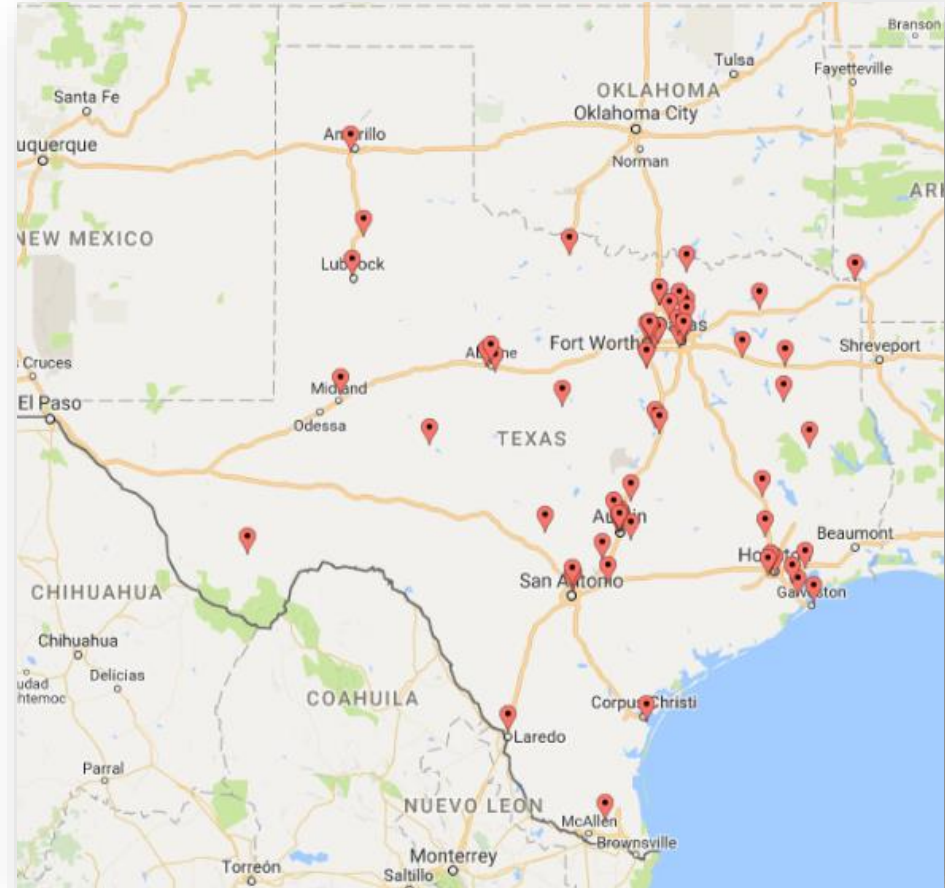
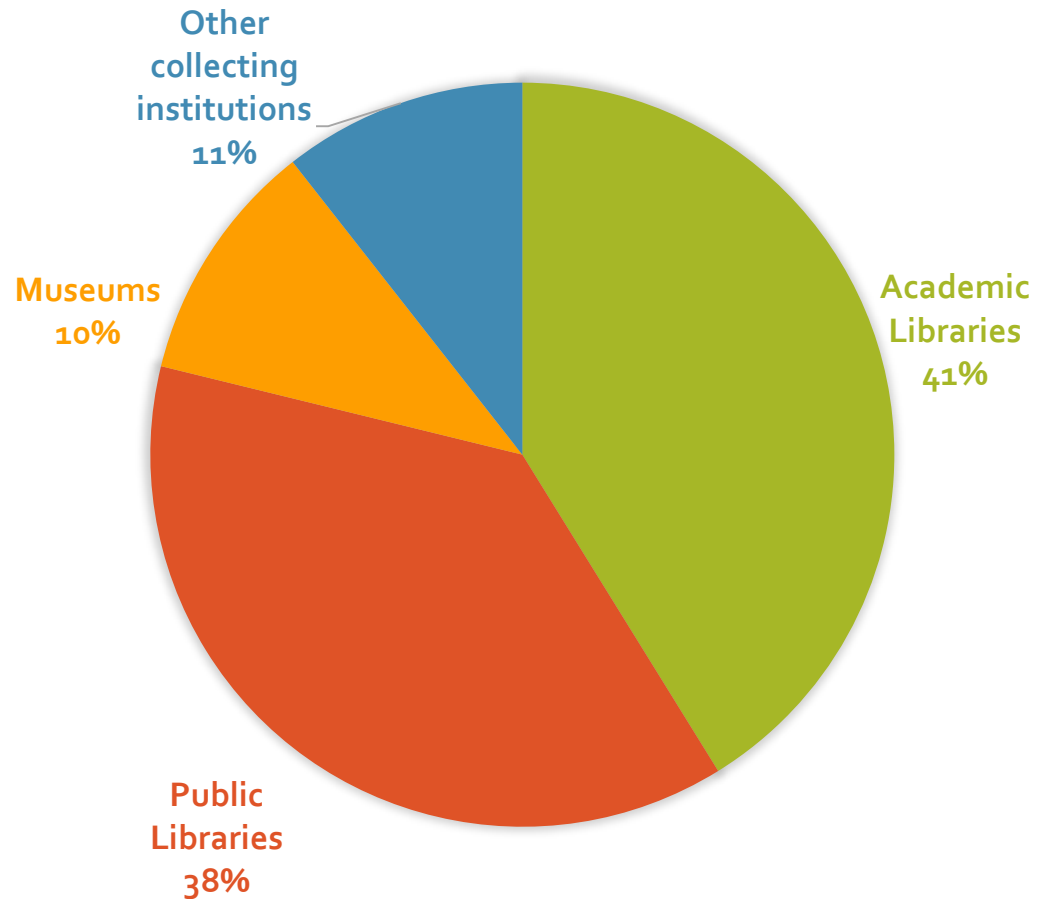
- ☐ Institutional Repository (IR) content/Scholarly research collections (such as electronic theses and dissertations, article pre-prints, research papers, or research data) (1)
- ☐ Newspaper collections (2)
- ☐ Audio/Video Collections (3)
- ☐ Government documents (4)
- ☐ Finding Aids/EADs (5)
- ☐ Other cultural heritage materials (6)
- ☐ Other (please specify) (7) \_\_\_\_\_

Q10 Tell us a little about the materials in your digital collections.

Q11 Please estimate the number of distinct collections of archival materials (not including Institutional Repository content or finding aids) are available in your institution digitally.

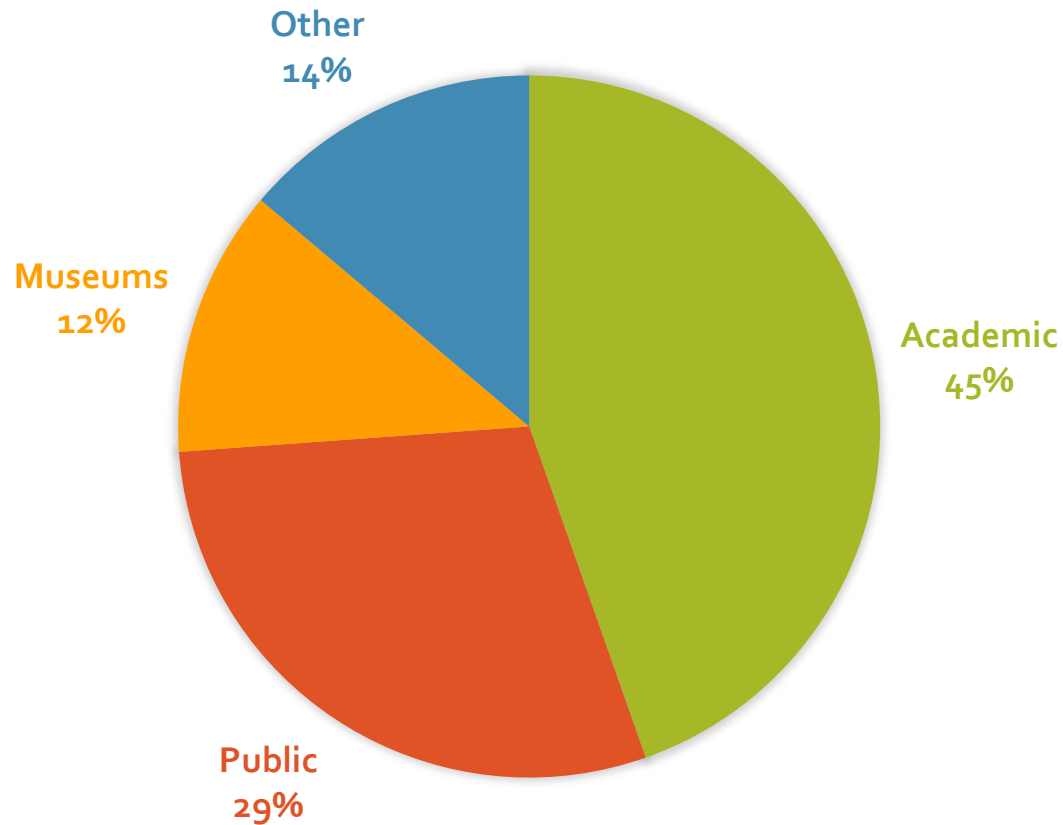
Survey results:

# Digital Collections Survey Responses



Survey results:

# Institutions with digital collections



Materials include...

- Photos
- Audio-visual materials
- Local & institutional history
- Finding Aids
- Newspapers
- Government documents
- Theses & dissertations
- Rare books and manuscripts
- Yearbooks

Survey results:

## Platforms & Metadata

- DSpace and CONTENTdm dominate
  - Others: Omeka, Islandora, Glifos, homegrown repositories, Polaris Fusion, database apps, etc.
- Majority use DC or Qualified DC (69%)
- Inconsistency in use of rights statements (academic libs more likely to include rights statements)

Survey results:

## Portal to Texas History

- A majority (58%) had materials in the Portal to Texas History.
- But only 8 (14%) had everything in the Portal.
- 78% have digital materials that aren't discoverable in DPLA

# Technology Development

- Working with 7 institutions to test the aggregator and find the pitfalls
- To come: develop search/browse front end for aggregated Texas collections



HARRY RANSOM  
HUMANITIES RESEARCH CENTER  
The University of Texas at Austin



TEXAS ARCHIVE  
OF THE MOVING IMAGE



RICE  
Fondren Library

# Metadata hubs survey

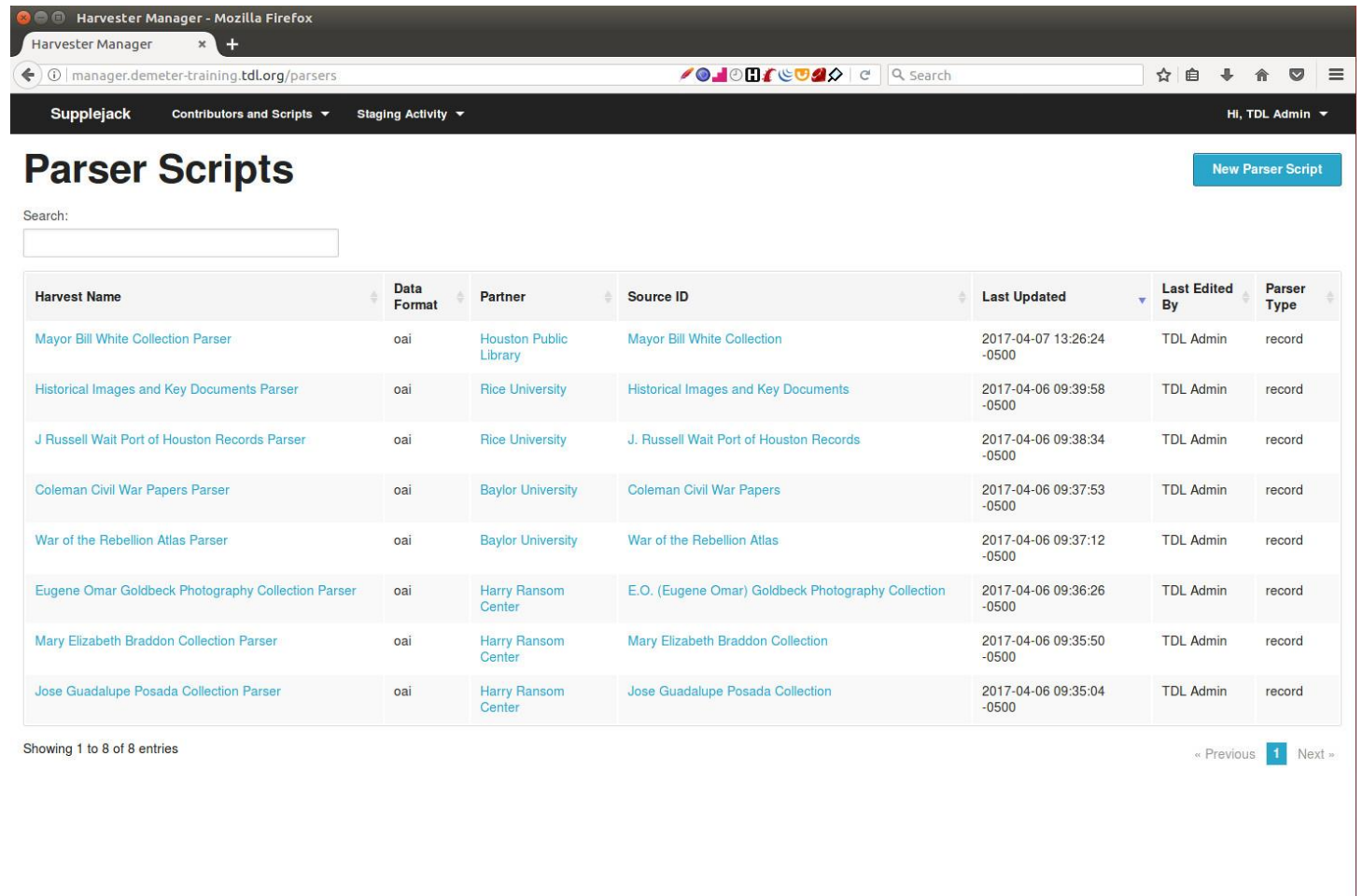
- Phone interviews with six metadata hubs
- Technology stacks: REPOX, custom Rails applications, Heidrun, others
- Diverse metadata sources and technology stacks, but large reliance on ContentDM and DSpace
- Most solutions depend on command line and lack GUI
- Cross-functional teams to handle all steps of the process

# Supplejack

- Developed at Digital Library of New Zealand
- Open source developer community
- Scalable to millions of records and hundreds of collections
- Technology stack
  - Ruby on Rails applications for the API, Harvest Manager and Worker
  - MongoDB to store metadata
  - Solr search index
  - Scheduled metadata harvesting Ruby libraries for cron jobs
  - Parallel metadata harvesting using multiple Ruby servers

# Collection parsers

- Four institutions
- Collections in ContentDM and DSpace
- OAI-PMH endpoints
- Compound objects



The screenshot shows the Harvester Manager web application in a Mozilla Firefox browser. The address bar displays the URL `manager.demeter-training.tdl.org/parsers`. The page title is "Harvester Manager". The navigation bar includes links for "Supplejack", "Contributors and Scripts", and "Staging Activity", along with a user profile "Hi, TDL Admin".

## Parser Scripts

Search:

Harvest Name	Data Format	Partner	Source ID	Last Updated	Last Edited By	Parser Type
<a href="#">Mayor Bill White Collection Parser</a>	oai	<a href="#">Houston Public Library</a>	<a href="#">Mayor Bill White Collection</a>	2017-04-07 13:26:24 -0500	TDL Admin	record
<a href="#">Historical Images and Key Documents Parser</a>	oai	<a href="#">Rice University</a>	<a href="#">Historical Images and Key Documents</a>	2017-04-06 09:39:58 -0500	TDL Admin	record
<a href="#">J Russell Wait Port of Houston Records Parser</a>	oai	<a href="#">Rice University</a>	<a href="#">J. Russell Wait Port of Houston Records</a>	2017-04-06 09:38:34 -0500	TDL Admin	record
<a href="#">Coleman Civil War Papers Parser</a>	oai	<a href="#">Baylor University</a>	<a href="#">Coleman Civil War Papers</a>	2017-04-06 09:37:53 -0500	TDL Admin	record
<a href="#">War of the Rebellion Atlas Parser</a>	oai	<a href="#">Baylor University</a>	<a href="#">War of the Rebellion Atlas</a>	2017-04-06 09:37:12 -0500	TDL Admin	record
<a href="#">Eugene Omar Goldbeck Photography Collection Parser</a>	oai	<a href="#">Harry Ransom Center</a>	<a href="#">E.O. (Eugene Omar) Goldbeck Photography Collection</a>	2017-04-06 09:36:26 -0500	TDL Admin	record
<a href="#">Mary Elizabeth Braddon Collection Parser</a>	oai	<a href="#">Harry Ransom Center</a>	<a href="#">Mary Elizabeth Braddon Collection</a>	2017-04-06 09:35:50 -0500	TDL Admin	record
<a href="#">Jose Guadalupe Posada Collection Parser</a>	oai	<a href="#">Harry Ransom Center</a>	<a href="#">Jose Guadalupe Posada Collection</a>	2017-04-06 09:35:04 -0500	TDL Admin	record

Showing 1 to 8 of 8 entries

« Previous 1 Next »

# Parser script editor

- Version control
- Metadata mapping
- XSLT paths
- Namespaces
- Conditional logic and functions in Ruby

The screenshot shows the Harvester Manager web interface in a Mozilla Firefox browser. The page title is "Mary Elizabeth Braddon Collection Parser" with a "record" tag. The main content area displays a Ruby script for parsing XML metadata. The script defines a class `MaryElizabethBraddonCollectionParser` that inherits from `SupplejackCommon::Oai::Base`. It sets the base URL to `http://hrc.contentdm.oclc.org/oai/oai.php` and the metadata prefix to `'oai_dc'`. It also defines namespaces for DC elements and OAI-DC. The script lists various attributes with their corresponding XPath expressions, such as `:content_partner`, `:primary_collection`, `:title`, `:description`, `:identifier`, `:creator`, `:publisher`, `:subject`, `:date`, `:type`, `:format`, `:language`, `:relation`, `:rights`, `:coverage`, and `:internal_identifier`. The `:internal_identifier` attribute uses a conditional XPath expression to select the last identifier. The script ends with `end`.

On the right side of the interface, there are several buttons and a history section. The buttons include "Preview", "Current version: initial commit", "Edit current version", "Rename Parser", "Change Data Source", "Delete Parser Script", and "Disable Full & Flush harvest mode". The "History" section shows a list of commits, including "initial commit" and "Applied parser template.", both dated "07 Apr 02:35 by TDL".

At the bottom of the interface, there is a "Message" input field and an "Update Parser Script" button.

# Harvest results preview

- Results in real time
- Raw data and metadata mapping
- API Record preview
- Usefulness depends on structure of feed results

The screenshot shows a web browser window titled "Harvester Manager - Mozilla Firefox". The address bar displays the URL "manager.demeter-training.tdl.org/parsers/58e6522f8784ef8ffc000006/edit". The main heading is "Mary Elizabeth Braddon Collection Parser" with a "record" button and a "Preview" button. Below the heading, there is a "Current version: initial commit" label. The central area is titled "Previewing records" and shows a status message "Status: Preview complete.". There are three tabs: "Source Data", "Harvested Attributes" (which is selected), and "API Record". The "Harvested Attributes" tab displays a JSON record structure. The "API Record" tab is also visible. The interface includes a sidebar with a code editor showing the parser's source code and a message box at the bottom left.

```
1 class MaryElizabethBraddonCollectionParser < SupplejackCommon::Dai::Base
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
end
```

Previewing records

Status: Preview complete.

Source Data Harvested Attributes API Record

< previous - Preview Record - next >

```
1 {
2   "priority": 0,
3   "match_concepts": null,
4   "content_partner": [
5     "Harry Ransom Center"
6   ],
7   "primary_collection": [
8     "Mary Elizabeth Braddon Collection"
9   ],
10  "title": [
11    "Diary: 1890"
12  ],
13  "description": [
14  ],
15  },
16  "identifier": [
17    "Manuscript Collection MS-4574",
18    "Series IV. Mary Elizabeth Braddon",
19    "Subseries A. Braddon Diaries",
20    "Box 26, Folder 1",
21    "http://hrc.contentdm.oclc.org/cdm/ref/collection/p15878coll53/id/162"
22  ],
23  "creator": [
24    "Braddon, M. E. (Mary Elizabeth), 1835-1915, author"
25  ],
26  "publisher": [
27    "Harry Ransom Center, The University of Texas at Austin"
28  ],
29  "subject": [
30    "Authors, English; English literature"
31  ],
32  "date": [
33    "1890-01-01T12:00:00.000Z"
34  ],
35  }
```

# Successes...

- Development and production Supplejack environments in AWS
- Custom metadata API
- Metadata harvesting of eight collections in ContentDM and DSpace

# and Challenges

- Environment setup and maintenance issues
- Handling metadata enrichments asynchronously from harvests
- Export and review of harvest results

# Technology Roadmap

- Manage the scheduling process for metadata harvests
- Separate server for dynamic image thumbnail generation
- OAI-PMH endpoint for re-harvesting of metadata
- Expand on collections in the pilot to include others that meet minimum criteria (e.g. OAI-PMH endpoint, metadata standards, etc.)
- Develop workflow and add staff to support regular metadata harvesting

# Long-term sustainability

- 1 FTE for training, outreach, and metadata coordination
- On-going software development support
- DPLA Hub Membership fee (coming soon)
- Governing body representing all hub constituents (publics, academics, museums & archives, etc.)

# Looking ahead

- Seeking additional grant funding for FY2017-18 to continue development and roll-out of aggregation service
- Work in phases, with collections meeting strict criteria included in phase 1
  - OAI endpoint
  - Inclusion of DPLA-required metadata elements, including rights statements (<https://dp.la/info/wp-content/uploads/2015/03/MA Pv4.pdf>)
  - Ability to provide metadata mapping to DPLA standard
  - Agreement to consider aggregated metadata as public domain



DIGITAL PUBLIC LIBRARY  
OF AMERICA

Texas DPLA Hub Governing Body/Bodies (HPL and others...)

**TEXAS AGGREGATION HUB & DISCOVERY  
SERVICE**



- Repository hosting

**PORTAL TO TEXAS HISTORY**

- Content hosting
- Digitization and metadata services



Contributing  
repositories

Texas Service Hub  
(Tentative structure)

# How can you participate?

- Tell us about your digital collections. ([info@tdl.org](mailto:info@tdl.org))
- Get involved with DPLA.
  - Community reps: <https://dp.la/info/get-involved/reps/>
  - Build something with the API: <https://dp.la/info/developers/>
  - Use DPLA resources: <https://dp.la/primary-source-sets>
- Contact info:
  - Kristi Park ([kristi.park@austin.utexas.edu](mailto:kristi.park@austin.utexas.edu))
  - Nick Woodward: [njw@austin.utexas.edu](mailto:njw@austin.utexas.edu)
  - General information: [info@tdl.org](mailto:info@tdl.org).