



# collections as data

thomas padilla

@thomasgpadilla

illustration by adam ferriss



*Always Already Computational: Collections as Data*  
2016-2018

thomas padilla, unlv

laurie allen, university of pennsylvania

stewart varner, university of pennsylvania

hannah frost, stanford university

sarah potvin, texas a&m university

elizabeth russey roke, emory university

THE  
ANDREW W.

**MELLON**  
FOUNDATION

*Collections as Data: Part to Whole*  
**2019-2021**

thomas padilla, unlv

hannah scates kettler, university of iowa

laurie allen, university of pennsylvania

stewart varner, university of pennsylvania

concept

initial motivations

implementing (imls)

implementing & supporting (mellon)

concept

initial motivations

implementing (imls)

implementing & supporting (mellon)

2014

# ~~library~~ collections as ~~humanities~~ data

Padilla, Thomas G., and Devin Higgins. 2014. “Library Collections as Humanities Data: The Facet Effect.” *Public Services Quarterly* 10 (4): 324–35. <http://bit.ly/2DrVcgs>

# 2014

Librarians can further enhance use of their digital collections by considering how thinking of them as ~~Humanities~~ data, and promoting them as such, can encourage uses beyond reading, viewing, and listening . . . efforts are ground in an understanding that data afford new opportunities for user interaction with library collections.

Padilla, Thomas G., and Devin Higgins. 2014. "Library Collections as Humanities Data: The Facet Effect." *Public Services Quarterly* 10 (4): 324–35. doi:10.1080/15228959.2014.963780



## DIGITAL PRESERVATION



Search this site

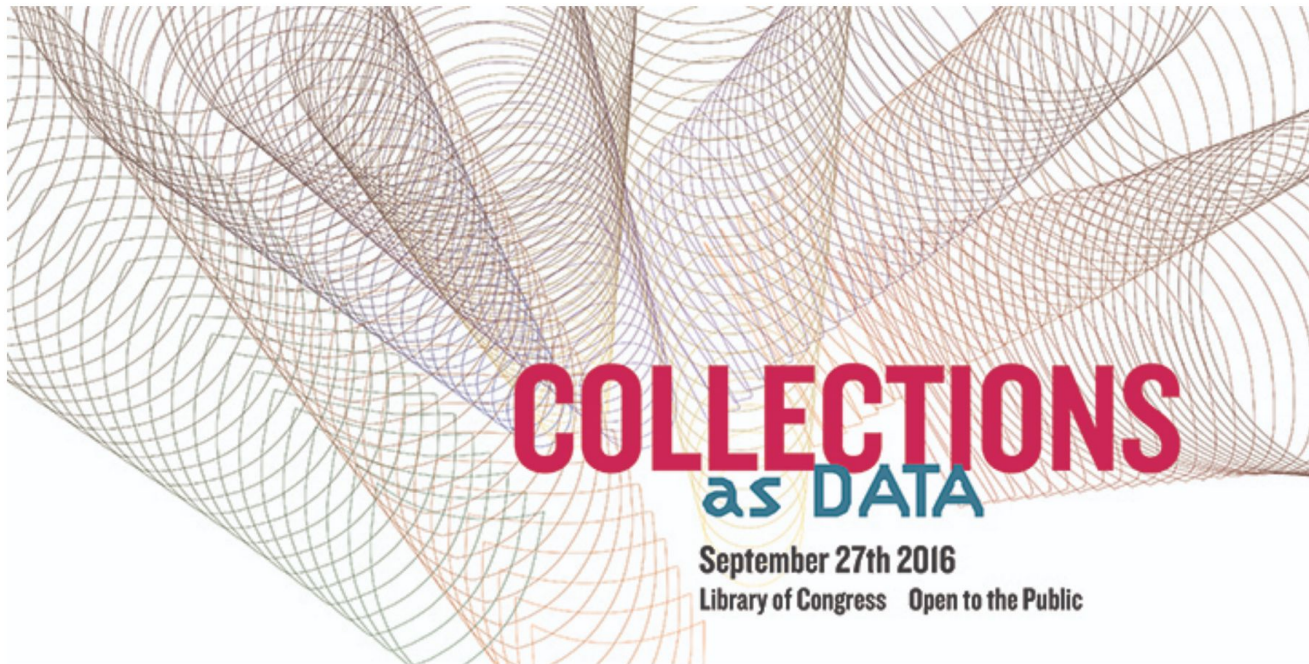
GO

- [Home](#)
- [About](#)
- [Meetings & Events](#)
- [Education & Training](#)
- [Contact Us](#)

### Resources

- > [Digital Formats Sustainability](#)
- > [Federal Agencies Digitization Guidelines Initiative](#)
- > [Library of Congress Recommended Format Specifications](#)

Print Subscribe Share/Save Give Feedback



## Collections as Data: Stewardship and Use Models to Enhance Access



# collections as data

... ordered information

... stored digitally

... amenable to computation



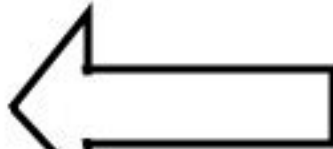


Button

Slider

Knob - Rotate

qualities that define  
possible use



Light  
Feedback



**procedural**

data affords a capacity for computational processing, *e.g. term frequency analysis, named entity extraction, and topic modeling*

**participatory**

data affords a capacity for enrichment by a diverse set of users, *e.g. crowdsourced transcription*

**encyclopedic**

data affords a capacity for expanded access, *e.g. parametric searching by granular features like line length, genre, author gender*

**spatial**

data affords a capacity for spatial characteristics to be surfaced, *e.g. place names can be geocoded and mapped*



# collections as data

- ... ordered information

- ... stored digitally

- ... amenable to computation

concept

initial motivations

implementing (imls)

implementing & supporting (mellon)





# Feeding America

Comments?

[MSU Libraries](#) » [DMC](#) » [Digital Collections](#) » [Feeding America](#) » [Browse](#) » [American Cookery](#)

Next Book



Hartford: Printed for Simeon Butler, Northampton, (1798)

Interest: The First American Cookbook & Northeast & The Great Ladies

[View Page Images](#)[Read Transcript](#)[View PDF](#)

**Digital Humanities**

About

Events

Humanities Data

Consultation

Guide

**Related Services**

Digital Scholarship Collaborative

## Feeding America: The Historic American Cookbook Project

```
<recipe class="soups" ethnicgroup="creole" region="south" occasion="other"><p><purpose align="center" size="smaller" placement="heading">SOUPE MAIGRE, WITHOUT MEAT, FOR LENT</purpose>
```

```
Melt half a pound of <ingredient>butter</ingredient> in a <implement>stewpan,</implement> put in six <ingredient>onions</ingredient> sliced; add two heads of <ingredient>celery</ingredient> cut small, one-half a head of <ingredient>white cabbage,</ingredient> and a bunch of chopped <ingredient>parsley;</ingredient> let them boil twenty minutes, then stir in three rolled <ingredient>crackers;</ingredient> pour in two quarts of boiling <ingredient>milk,</ingredient> or <ingredient>milk</ingredient> and <ingredient>water;</ingredient> let this boil up gently for half an hour, and just before serving stir in two well-beaten <ingredient>eggs.</ingredient></p></recipe>
```



## Download Data

[The Feeding America Dataset](#)

## Description

The "Feeding America: The Historic American Cookbook" dataset contains transcribed and encoded text from 76 influential American cookbooks held by MSU Libraries Special Collections. Features encoded within the text include but are not limited to recipes, types of recipes, cooking implements, and ingredients. The 76 texts were chosen among more than 7000 cookbooks that MSU Libraries holds as representative of periods and themes in American cookbook history spanning the late 18th to early 20th century.

[lib.msu.edu/feedingamericadata](http://lib.msu.edu/feedingamericadata)



UNC  
UNIVERSITY LIBRARIES



DIGITAL PUBLIC LIBRARY  
OF AMERICA



Penn  
Libraries



COOPER  
HEWITT



# Cultural Heritage & Library Collections as Data and their Role in Digital Humanities Infrastructure

◆ eResearch in Humanities and Social Sciences

Dr T and encouraging reuse of collections that support computationally-driven research.” [2] One of the drivers for this initiative is the

<sup>1</sup>Uni perception that, as Miriam Posner argues, “Libraries and archives

<sup>2</sup>Dez [and museums] are increasingly making their materials available

<sup>3</sup>Dez online, but, as a general rule, these materials aren’t of much use  
**AB!** for computational purposes.” [3]

The  
form  
For t  
the c  
toge

At th and Instead, as Thomas Padilla has summarized the project, a  
“collections as data” imperative can benefit research communities

providing access to, and encouraging reuse of collections that support computationally-driven research.” [2] One of the drivers for this initiative is the perception that, as Miriam Posner argues, “Libraries and archives [and museums] are increasingly making their materials available online, but, as a rule, these materials aren’t of much use for computational purposes.” [3]

and dig  
human  
als that t  
ity as pr

ry of Co  
cribing,



# FOREIGN RELATIONS OF THE UNITED STATES

CITATION URL: <http://digital.library.wisc.edu/1711.dl/FRUS>

 BROWSE THE COLLECTION

 SEARCH THE FULL TEXT

[BROWSE](#) > FOREIGN RELATIONS OF THE U.S.

Foreign Relations  
of the  
United States  
1952-1954  
Volume II  
National Security  
Affairs

(in two parts)  
Part 1



Editor in Chief  
William Z. Slattery

Editors  
Lester A. Sloan  
Nash H. Petersen

United States  
Government Printing Office  
Washington : 1954

This digital facsimile of *Foreign Relations of the United States* is a project of the University of Wisconsin-Madison Libraries in collaboration with the University of Illinois at Chicago Libraries. This is a nearly complete run from 1861-1960 with missing volumes being added as they can be acquired and processed. If your library is interested in donating material for this project, please contact the [University of Wisconsin Digital Collections Center](#).


Special thanks are owed to the University of Chicago Libraries for contributing FRUS volumes for digitization. Thanks also to these other contributing libraries: Southwestern University (Los Angeles) Library, the University of Connecticut's Homer Babbidge Library, Georgia Southwestern State University's James Earl Carter Library, University of Indiana-Bloomington Libraries, Brigham Young University's Howard W. Hunter Law Library, Columbia University Libraries, Des Moines Public Library, Jefferson County (Colorado) Public Library, Michigan State University Libraries, Ouachita Baptist University's Riley-Hickingbotham Library, University of Michigan-Flint's Thompson Library, University of Minnesota, University of Mississippi's J. D. Williams Library, University of South Alabama's

Foreign Relations of  
the United States,



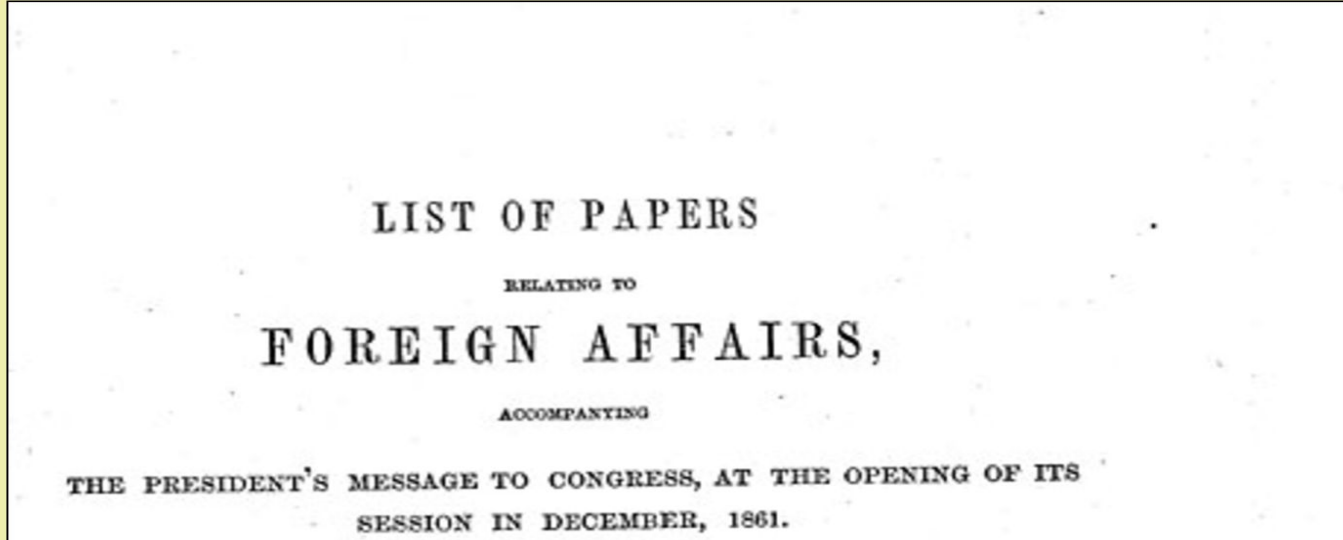
**United States Department of State / *Message of the President of the United States to the two houses of Congress, at the commencement of the second session of the thirty-seventh congress***  
**(1861)**

[Contents] List of papers relating to foreign affairs, accompanying the president's message to Congress, at the opening of its session in December, 1861, pp. [21]-29 ff.

 [PDF \(2.4 MB\)](#)

---

**Page [21]**





thomasgpadilla Update frus\_section\_pdf\_urls.py

48c1b58 on Aug 10, 2016

1 contributor

36 lines (26 sloc) | 1.22 KB

Raw

Blame

History



```
1  from urllib2 import urlopen
2  from bs4 import BeautifulSoup
3  import requests
4
5  #point to FRUS volumes
6  html = urlopen("http://digicoll.library.wisc.edu/cgi-bin/FRUS/FRUS-idx?type=browse&scope=FRUS.FRUS1")
7  soup = BeautifulSoup(html, 'html.parser')
8
9  with open("frus_section_pdf_urls.txt", "w") as file:
10
11  #scrape FRUS volume relative URLs
12      for links in soup.find_all("p", {"class":"isshead"}):
13
14  #prepend relative URLs with base URL
15      for link in links.find_all('a'):
16          volumeurl = 'http://digicoll.library.wisc.edu/' + link.get('href')
17
18          volumeurls = requests.get (volumeurl)
19          volumesoup=BeautifulSoup(volumeurls.content, 'html.parser')
20
21  #scrape FRUS volume section relative URLs
22      for urls in volumesoup.find_all("p", {"class":"cntsitem"}):
23
24  #prepend relative URLs with base URL
25      for url in urls.find_all('a'):
26          pdfurl = 'http://digicoll.library.wisc.edu/' + url.get('href')
```

**There is a gulf between the research processes of academic researchers and the curatorial processes of the cultural institutions which hold most of the source materials**—these institutions have their own ways of organising and describing source materials which may be quite different from the information produced by the research process

Toby Burrows

*Sharing humanities data for e-research: conceptual and technical issues*

**all1700-99.tar.gz** (7.06 MB)

**all1800-49.tar.gz** (42.23 MB)

**all1850-74.tar.gz** (40.19 MB)

**all1875-99.tar.gz** (82.07 MB)

The genre predictions were produced by an ensemble of regularized logistic classifiers, and are intended to support research that explores broad trends in literary history. Since volumes usually contain multiple genres, page-level metadata is necessary to create machine-readable collections in a particular genre.

Predictions are included as JSON objects in separate files, one for each volume. The tar.gz files prefixed with "all" include all 854,476 volumes, divided by date. The tar.gz files named for genres contain subsets of volumes that have been filtered to achieve greater than 97% precision in that particular genre. Specifically, they include 18,111 vols containing drama, 102,349 vols containing fiction, and 61,286 vols containing poetry. These datasets were filtered both with confidence metrics created by a logistic model and by manual editing. Ringers.csv is a list of volumes that we had to manually remove; scholars who select their own datasets from the larger collection (of files beginning "all") may also want to consider filtering out these tricky cases.

- **Miriam Posner:** to what extent is information about . . . collection provenance, processing, and method of presentation available to the user?
- **Victoria Stodden:** to what extent are data and the code that generates data available to the user?
- **Roopika Risam:** to what extent are the motivations driving all of the above available to the user?

... an imperative present



## Robert Creeley Email Archives

Date Range: January 1, 1904 to June 20, 2009

Messages: 31278 incoming, 18651 outgoing.



Correspondents (24353)



Persons (57556)



Organizations (42300)



Locations (22131)



# DocNow

DocNow is a tool and a community developed around supporting the ethical collection, use, and preservation of social media content.

Photo by [Robert Cohen](#)

concept

initial motivations

implementing (imls)

implementing & supporting (mellon)



*Always Already Computational: Collections as Data*  
2016-2018

thomas padilla, unlv

laurie allen, university of pennsylvania

stewart varner, university of pennsylvania

hannah frost, stanford university

sarah potvin, texas a&m university

elizabeth russey roke, emory university

American Historical Association; American Philosophical Society; British Library; Carnegie Museum of Art; Compute Canada; Cornell University; Deutsches Klimarechenzentrum; Digital Library Federation; Digital Public Library of America; Emory University; Getty Research Institute; Hathitrust Research Center; Haverford College; Indiana University; Indiana University-Purdue University; Internet Archive; James Madison University; Koninklijke Bibliotheek; Library of Congress; Max Planck Computing & Data Facility; McGill University; Michigan State University; Massachusetts Institute of Technology; Museum of Modern Art; National University of Singapore; New York Public Library; New York University; Northeastern University; Open Knowledge Foundation; Penn State University; Stanford University; Swarthmore College; Texas A&M University; Tufts University; University College London; University of British Columbia; University of California Santa Barbara; University of California Berkeley; University of California Los Angeles; University of Canberra; University of Delaware; University of Graz; University of Houston; University of Illinois at Urbana Champaign; University of Maryland; University of Miami; University of Minnesota; University of North Carolina at Chapel Hill; University of Pennsylvania; University of Toronto; University of Utah; University of Texas Austin .....

# 2016

predominant digital collection development focuses on replicating traditional ways of interacting with objects in a digital space. this approach does not meet the needs of the researcher, the student, the journalist, and others who would like to [work with] computational methods and tools to treat ... collections as data.

Thomas Padilla, Laurie Allen, Stewart Varner, Sarah Potvin, Elizabeth Russey Roke, Hannah Frost. 2016. National Leadership Grant. Institute for Museum and Library Services.

tl:dr

how do we make  
collections **more useful?**



what **gaps** does this  
question surface?

what **opportunities** does  
this question **help us see?**

... for me

... for you

... for **the people** that  
we aspire to collaborate  
with, partner with, and  
serve

creation of **a series of resources** that:

1. **document** extant experience
2. **share** lessons learned
3. **suggest** a range of paths into the work

**activities,** deliverables, impacts

# by the numbers

- 2 years
- 2 National Forums
- 17 f2f & virtual conference engagements
- 1 full day international pre-conference
- 2 webinars
- 2 conference workshops
- 1 week-long DH institute course w/ British Library
- 382 people join the CAD [google group](#)
- 59 people join the CAD [zotero group](#)

**7 primary deliverables!**







- Santa Barbara Statement on Collections as Data
- Facets
- Personas
- Position Statements
- Methods
- 50 Things
- White paper (to be released)

# The Santa Barbara Statement on Collections as Data

*Version 2*

*The Santa Barbara Statement on Collections as Data was written by the Institute of Museum and Library Services supported Always Already Computational: Collections as Data project team. The first version is based on the collaborative work of participants at the first Collections as Data National Forum (UC Santa Barbara, March 1-3 2017). After its release, the team gathered comments from the Hypothesis web annotation tool and sought additional feedback across a series of conversations and workshops (April 2017 - April 2018). The current version of the statement was revised based on that community feedback, especially the close, directed feedback provided by workshop participants at the Digital Library Federation Forum 2017.*



# Principles

1. **Collections as data development aims to encourage computational use of digitized and born digital collections.** By conceiving of, packaging, and making collections available as data, cultural heritage institutions work to expand the set of possible opportunities for engaging with collections.
2. **Collections as data stewards are guided by ongoing ethical commitments.** These commitments work against historic and contemporary inequities represented in collection scope, description, access, and use. Commitments should be formally documented and made publicly available. Commitment details will vary across communities served by collections but will share common cause in seeking to address the needs of the vulnerable. Collection stewards aim to respect the rights and needs of the communities who create content that constitute collections, those who are represented in collections, as well as the communities that use them.
3. **Collections as data stewards aim to lower barriers to use.** A range of accessible instructional materials and documentation should be developed to support collections as data use. These materials should be scoped to varying levels of technical expertise. Materials should also be scoped to a range of disciplinary, professional, creative, artistic, and educational contexts. Furthermore the community should be motivated and encouraged to build and share tools and infrastructure to facilitate use of collections as data.
4. **Collections as data designed for everyone serve no one.** Specific needs inform collections as data development. These needs may be commonly held by particular user communities. Rather than assuming these needs or imagining these communities, stewards should be intentional about who their collections are designed for, work to lower the barriers to use for the people in those communities, and continue to assess these needs over time. Where resources permit, multiple approaches to data development and access are encouraged.

## Collections as Data Facets

facet \ 'fa-sət\ : one side of something many-sided

=====

*Collections as Data Facets* document collections as data implementations.

**An implementation consists of the people,  
services, practices, technologies, and infrastructure that aim to  
encourage computational use of cultural heritage collections.**

[collectionsasdata.github.io/facets](https://collectionsasdata.github.io/facets)

1. Why do it
2. Making the Case
3. How you did it
4. Share the docs
5. Understanding use
6. Who supports use
7. Things people should know
8. What's next

### **Facet 1** - MIT Libraries Text and Data Mining

- Richard Rodgers, Massachusetts Institute of Technology Libraries

### **Facet 2** - Carnegie Museum of Art Collection Data

- David Newbury, Carnegie Museum of Art and Daniel Fowler, Open Knowledge International

### **Facet 3** - CalCOFI Hydrobiological Survey of Monterey Bay

- Amanda Whitmire, Stanford University Libraries

### **Facet 4** - American Philosophical Society Open Data Projects

- Scott Ziegler, American Philosophical Society

### **Facet 5** - OPENN

- Dot Porter, University of Pennsylvania Libraries

### **Facet 6** - Chronicling America

- Deborah Thomas, Nathan Yarasavage, and Robin Butterhof, Library of Congress


### **Facet 7** - La Gaceta de la Habana

- Paige Morgan, Elliot Williams, and Laura Capell, University of Miami Libraries



## Collections as Data Personas

persona /ˌpərˈsōnə/: an archetypal user

Faculty	Motivations	Specific Goals
	<ul style="list-style-type: none"><li>• Desire to uncover previously unknown information and create new insight</li><li>• Need to earn the respect and trust of colleagues by producing high quality, interpretive scholarship based on solid research</li><li>• Ambition to push the boundaries of traditional methods and take advantage of new research opportunities</li><li>• Eagerness to design new assignments that will harness their students' interest in technology and teach them how to be effective and critical digital citizens</li><li>• Interest in helping to develop better collections</li></ul>	<ul style="list-style-type: none"><li>• Locate and download data that is trustworthy and appropriately formatted to work with their preferred software</li><li>• Develop assignments for a new course with the Library's recently appointed Digital Humanities Specialist</li><li>• Submit request to the IT department for 1 Terabyte of online storage for data projects so it is easy to review student work and so that it is easy to share with peers during peer review or after publication</li></ul>
<p><b>Dr. Kylie Yu</b> is an associate professor of US history at a small liberal arts college in the midwest. Their research is focused on representations of immigrants in the media in the 19th century and they teach the survey of US history, 1865 - Present, as well as upper level courses for history majors on immigration, identity, and urban life. They are not technophobic, but they have very little formal technical training.</p>	<p>"I am very excited about the potential for digital methods to allow me to ask different questions than I am used to, but I don't think I have the tech skills to create useful datasets so I can begin experimenting."</p> <p>"I was initially very excited about the digitized newspaper collections here, but the software I wanted to use only works on plain text, not PDFs, and I simply don't have time to download each page individually. I was also disappointed in the poor quality of the OCR text. Then I discovered the enhanced availability of computationally-accessible newspaper content provided by the Library of Congress and it blew my mind to consider all the research possibilities. This is going to be fun!"</p>	

# 50 things you can do

---



Over the course of our two-year project, members of the project team often heard a question along the lines of: “We’ve read the *Santa Barbara Statement*, *the Facets*, *the Personas*. We’re subscribed to the *google group*. But where do we start? How do we move collections as data forward at our own institution?”





- ... taken up as a strategic priority within the University of California's Shared Content Leadership Group's *Plans & Priorities for 2017/2018 Based on the University of California Library Collection: Content for the 21st Century and Beyond*
- ... incorporated as a feature of the OCLC *Research and Learning Agenda for Archives, Special, and Distinctive Collections in Research Libraries*
- ... inform the creation of postdoctoral positions like the British National Archives' FTNA Postdoctoral Research Fellowship, focused on unlocking, "archival collections as data"
- ... presented as a component of the Digital Library Federation eResearch Network
- ... inform the Software Preservation Network's outreach
- ... inform the development of a week-long collections as data course at the Humanities Intensive Learning and Teaching Institute
- ... inspire reading groups, international hackathons, workshops, and conference sessions that span disciplinary, library, archives, and museum communities.

# Have you used it?

Always Already Computational: Collections as Data is drawing to a close!

<https://collectionsasdata.github.io/>

If you've used anything from the project we'd love to hear about it. Have you used the Santa Barbara Statement on Collections as Data? The personas? The facets? Has the project helped support any actions in your local, regional, or larger professional context? For example, a change in workflows, an influence on strategic directions, contributing to external funding effort, development of workshops, (re)defining roles and responsibilities?

Any and all responses are welcome. We intend to share responses publicly. If you'd prefer that your name not be shared with your response please indicate in the option below.

Many thanks,

Thomas Padilla, University of Nevada Las Vegas

Laurie Allen, University of Pennsylvania

Stewart Varner, University of Pennsylvania

Sarah Potvin, Texas A&M University

Elizabeth Russey Roke, Emory University

Hannah Frost, Stanford University

IMLS LG-73-16-0096-16

[bit.ly/haveyouusedit](https://bit.ly/haveyouusedit)

(1) I'm leading data curation work package in a national research and data infrastructure project for humanities, arts and social sciences . . . The facets have ... informed the development of a data curation framework for data sharing and interoperability across multiple platforms (discovery, access, research and archiving).

Ingrid Mason,  
Australia's Academic and Research Network (AARNet)

I use it as a communication and teaching tool about what digital archives are, why they are important, how they play a role in the civic literacy, social justice, and community engagement aspects of our College's mission, and what groundwork must be laid infrastructurally in order to support the building of collections as data.

Charlotte Nunes,  
Lafayette College Libraries



- Santa Barbara Statement on Collections as Data
- Facets
- Personas
- Position Statements
- Methods
- 50 Things
- White paper (to be released)

concept

initial motivations

implementing (imls)

implementing & supporting (mellon)





*Collections as Data: Part to Whole*  
**2019-2021**

thomas padilla, unlv  
hannah scates kettler, university of iowa  
laurie allen, university of pennsylvania  
stewart varner, university of pennsylvania



**Collections as Data: Part to Whole** aims to foster the development of broadly viable models that support implementation *and* use of **collections as data**.

**Collections as Data: Part to Whole** will  
regrant **\$600,000** to **12** projects across **2**  
cohorts.





cohorts will create:

1. **implementation** models
2. **organizational** models
3. **new collections that speak to underrepresented histories**

cohorts will take part in:

1. **team lead institute**
2. **public summative forum**



## **Cohort 1 Projects (1/2019-4/2020)**

---

### **Collections as Data: Redefining Creators, Users, and Stewards of the Charles “Teenie” Harris Photographic Archival Collection**

Carnegie Museum of Art

Dominique Luster, Charlene Foggie-Barnett, Ed Motznik, Samantha Ticknor

*This project seeks to build upon the rich history of the Teenie Harris archival collection and develop new opportunities for computer-generated creation and computational manipulation of collection metadata that is both produced and used by the African American community and the Carnegie Museum of Art. This project aims to develop and document the service and use capabilities and limitations of machine learning, text parsing, and computer vision technologies to make meaningful contributions to archival metadata. The public facing deliverables will combine the notion of creators and users of the Harris data and will result in a suite of web-based in-gallery interactives that have the functionality of engagement with and contribution to the collection as data.*

## **Uncovering Health History: Transcribing and Publishing Early Twentieth-Century Tuberculosis Patient Records as Data**

University of Denver

Kim Pham, Kevin Clair, Jack Maness, Jeanne Abrams, Fernando Reyes, Jeff Rynhart, Alice Tarrant

*This project will use handwritten text recognition (HTR) to create transcriptions of records that have been unavailable through traditional OCR processes from the Jewish Consumptives' Relief Society, a tuberculosis sanatorium located in Denver from 1904 to 1954. These records represent a valuable archive of primary source materials regarding the treatment history of tuberculosis in the early 20th century and the history of primarily Jewish and Eastern European immigration to Denver during the same time period. We intend to develop capacity in services and infrastructure to support and use HTR technologies in our regular workflow to produce collections as data and to contribute to the emerging HTR technology ecosystem.*

# timeline

- Cohort 1 initiated (1/2019)
- Cohort 2 CFP opens (8/2019)
- Cohort 2 CFP closed (10/2019)
- Cohort 2 initiated (1/2020)
- Cohort 1 Public Forum (1/2020)
- Cohort 1 Models & Collection Release (4/2020)

# questions about cfp

- face to face
  - spring CNI 2019 (thomas, hannah, laurie)
  - ACRL 2019 (thomas, hannah)
  - csvconf (thomas)
- email/video chat/phone
  - [thomas.padilla@unlv.edu](mailto:thomas.padilla@unlv.edu); [hannah-s-kettler@uiowa.edu](mailto:hannah-s-kettler@uiowa.edu);  
[laallen@upenn.edu](mailto:laallen@upenn.edu); [svarner@upenn.edu](mailto:svarner@upenn.edu)



## Welcome to Labs!

A place to encourage innovation with Library of Congress digital collections.



### Experiments

Check out tools, art, applications, and visualizations we've made with our collections.



### LC for Robots

Find APIs, bulk downloads, and tutorials for the Library's digital collections.



### Events

See what's coming up, and read about past events.

## CDH's Hackathon Connects Computational Researchers From Across Campus

Suzanne S. Roth

January 30, 2019

The Center for Digital Humanities (CDH) recently hosted an all-day Hackathon where developers across campus joined together in teams to explore library collection data with a digital humanities lens by applying computationally driven methods to analyze the archives in creative ways. Participants included more than 20 developers from CDH, the Princeton University Library (PUL) and the Princeton Institute for Computational Science and Engineering/Research Computing (PICSciE/RC).



ARLIS/NA  
47th Annual Conference  
March 26 - 30, 2019  
Salt Lake City

# ARTS:IN/SITE

INSIGHT  INCITE  IN SIGHT

PRELIMINARY SCHEDULE - SUBJECT TO CHANGE

[Schedule ▼](#) [Speakers](#) [Sponsors](#) [Exhibitors](#) [Attendees](#)

**Wednesday, March 27 • 4:10pm - 5:30pm**

Getting Started with Collections as Data



# MW19 | Boston

Boston, Massachusetts, USA, April 2-6, 2019

[MuseWeb](#)

[About](#) ▾

[Attending](#) ▾

[Presenting](#) ▾

[Exhibiting](#) ▾

[Conference](#) ▾

## Cultural Collections as Data: Aiming for digital data literacy and tool development

Other

[Liz Neely](#), Georgia O'Keeffe Museum, USA, [Anne Luther](#), Center for Data Arts, The New School, USA, [Chad Weinard](#), Williams College Museum of Art, USA