



# Public Humanities Data

Micah Vandegrift  
@micahvandegrift  
[bit.ly/2CreDZM](https://bit.ly/2CreDZM)





**IMMERSIVE**Scholar



# Liss LaFleur

Join us for a talk by creative resident and feminist artist  
Liss LaFleur, exploring #MeToo through stained glass.

**Thursday, September 13, 9:30 a.m.**  
**Hunt Library, Teaching and Visualization Lab**

LaFleur's residence is funded through the Andrew W. Mellon Foundation.

THE  
ANDREW W.  
**MELLON**  
FOUNDATION



# Digital Scholarship Workshop Series

SEP

14

Introduction to XML and Text Encoding

September 14

1:00 PM to 3:00 PM

AT THE

ITTC Lab 2, D. H. Hill Library

OCT

24

Storytelling with GIS

October 24

1:00 PM to 3:00 PM

AT THE

ITTC Lab 2, D. H. Hill Library

OCT

26

Getting Started with Building Digital Exhibits in Omeka

October 26

10:00 AM to 12:00 PM

AT THE

ITTC Lab 2, D. H. Hill Library

[go.ncsu.edu/dsworkshops](http://go.ncsu.edu/dsworkshops)

Compiled from ideas and words shared by these deep thinkers and colleagues:

Alex Gil

Miriam Posner

Trevor Muñoz

Brandon Locke

Julia Flanders

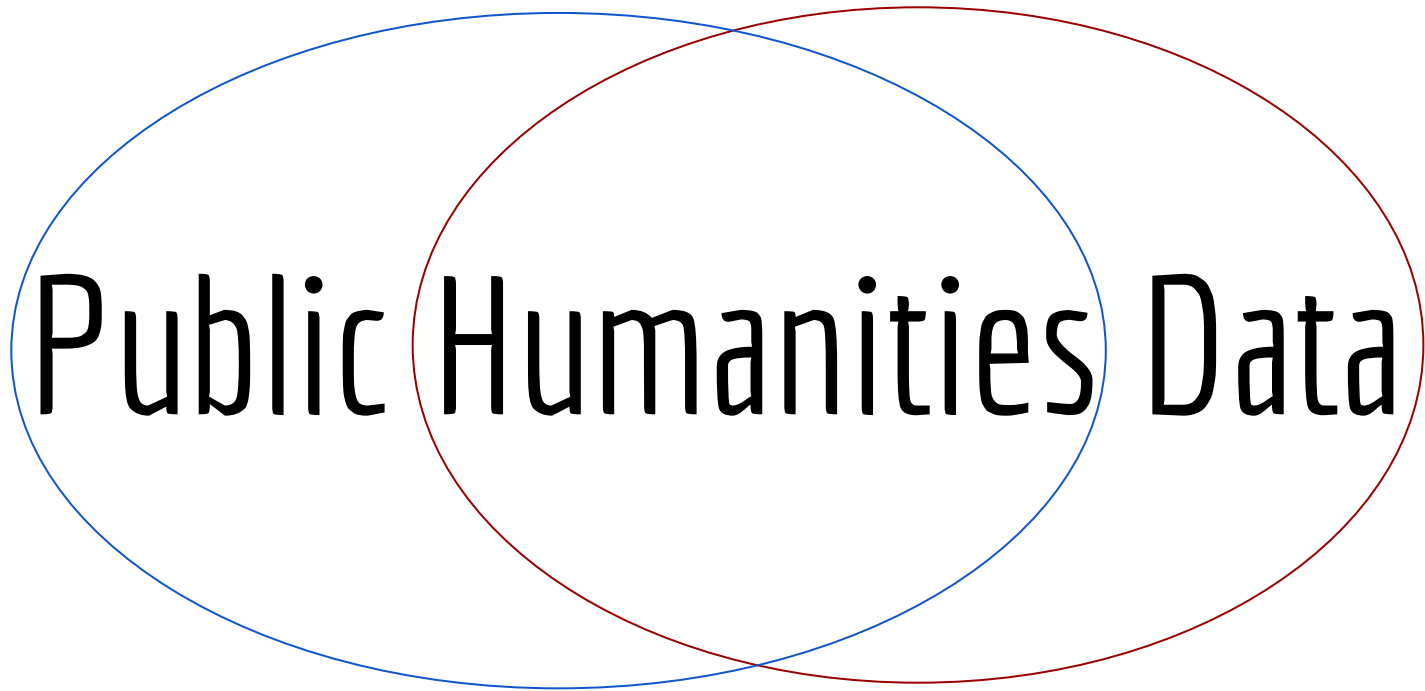
Thomas Padilla

Cynthia Hudson-Vitale

Katie Rawson

Please google their work and cite it often





# Why public humanities data?



**Jessica Pickett**  
@pickettiessica

Follow

This is a great example of the need for the digital & public humanities to engage more closely with big data / machine learning (and vice versa).

Frequently, some groups are over-represented and others are under-represented. More than 42% of Instagram photos, which have been taken in computer vision, come from the United States<sup>1</sup>, more to only 4% of the world's population. By contrast, China and India together contribute just 2% of Instagram data, even though these countries represent 36% of the world's population. This lack of generalisability partly explains why computer vision algorithms find a photograph of a traditional US bride dressed in white as 'bride', 'brun', 'woman', 'wedding', but a photograph of a North Indian bride as 'performance art' and

Rachel Thomas @math\_rachel

computer vision algorithms trained on ImageNet label a photo of a traditional US bride dressed in white as 'bride', 'dress', 'woman', 'wedding', but label a photo of a North Indian bride as 'performance art' and 'costume'

11:53 PM - 23 Jul 2018

1 Retweet 5 Likes



↕ 1



5

the “data” turn, akin to [spatial.scholarslab.org](https://spatial.scholarslab.org)

# What are humanities data?



Document



Dataset



Image



Sensor



Ontology



Geospatial



Project



Collection



# What are humanities data?

Working definition:  
machine-readable artifacts of the  
human condition through time





# Where are humanities data?



News About Using tDAR **UPLOAD**

## What can you dig up?

The Digital Archaeological Record (tDAR) is your online archive for archaeological information.



advanced

**SIGN UP**

**LOG IN**



### Access & Use

Broadening the access to archaeological data through simple search and browse functionality.

**LEARN MORE**



### Preservation

Dedicated to ensuring long-term preservation of digital archaeological data.

**LEARN MORE**



### Upload Resources

Contribute documents, data sets, images, and other critical archaeological materials.

**LEARN MORE**



### Who Uses tDAR

Researchers like you. Uncover knowledge of the past, and preserve and protect archaeological resources.

**LEARN MORE**

core.tdar.org



Search

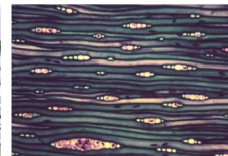
Browse by Topic > Women in Science

## Women in Science



### Education

Learn about the history of women's formal study in the sciences. Delve into early arguments on both sides of the question of women's education, including the work of advocates Mary Astell, Anna Brackett, Alice Freeman Palmer, and the Association of Collegiate Alumnae. Explore programs such as science fairs that encourage an interest in science and browse the history of women in higher education at both women's colleges and coeducational institutions.



### Botany

Explore the evolution of women's involvement in the study of botany, from the emergence of botanical illustration and plant collecting as acceptable pursuits for women, through formal education, and pioneering work in the field and the lab. Unearth the lush illustrations of Maria Sibylla Merian, writings on California botany by Alice Eastwood, letters and photographs from Mary Agnes Chase's expedition to Brazil, and microscopic images from Katherine Esau's lab.




### Medicine

Discover women's impact in the field of medicine, as doctors, nurses, pharmacists, and researchers. Learn about groundbreaking doctors such as Elizabeth Blackwell, Rebecca Lee Crumpler, Susan La Flesche Picotte, and others. Explore key achievements such as the founding of the Red Cross, the development of a Nobel prize-winning technique for blood screening, and the invention of a standardized way to assess the health of newborns.

dp.la

# Where are humanities data?

HathiTrust Research Center | JSTOR Data for Research | EEBO-TCP

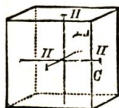


**HathiTrust Research Center Analytics**

Supports large-scale computational analysis of the works in the HathiTrust Digital Library to facilitate non-profit and educational research.

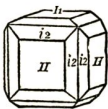
New updates are live in HTRC Analytics! When you log in, check out the updated suite of algorithms. You'll also find new, in-browser access to HTRC Data Capsules. Give them spin and let us know what you think!

## Featured Services



### Extracted Features

An unrestricted dataset of metadata and word counts for each page in the HathiTrust Digital Library. Download and explore on your own machine.



### Text Analysis Algorithms


Web-based, click-and-run tools that perform computational text analysis on worksheets, which are user-created collections of volumes. No programming required.



### Data Capsules

Secure virtual environments for non-consumptive text analysis, where researchers can implement their own data analysis and visualization tools.

[analytics.hathitrust.org](http://analytics.hathitrust.org)



Advanced Search | Browse | Login to My Account | Register | About | Support

## JSTOR Data for Research

Define and request datasets for content on JSTOR, or download a sample dataset for teaching text mining techniques.

### Create a Dataset

Data for Research (DFR) provides datasets of content on JSTOR for use in research and teaching. Researchers may use DFR to define and submit their desired dataset to be automatically processed. Data available through the service includes metadata, n-grams, and word counts for most articles and book chapters, and for all research reports and pamphlets on JSTOR. Datasets are produced at no cost to researchers and may include data for up to 25,000 documents.

[Create a Dataset](#) [How to create a dataset](#)

### Large & Full-text Datasets

Large and full-text datasets are provided upon request and require an agreement about the use of the data.

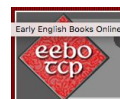
[Request a Dataset](#) [About Dataset Services](#)

### Sample Datasets

Download a sample dataset for exploration or use in teaching text mining techniques and methods.


[Explore Sample Datasets](#)

[jstor.org/df](http://jstor.org/df)



Early English Books Online

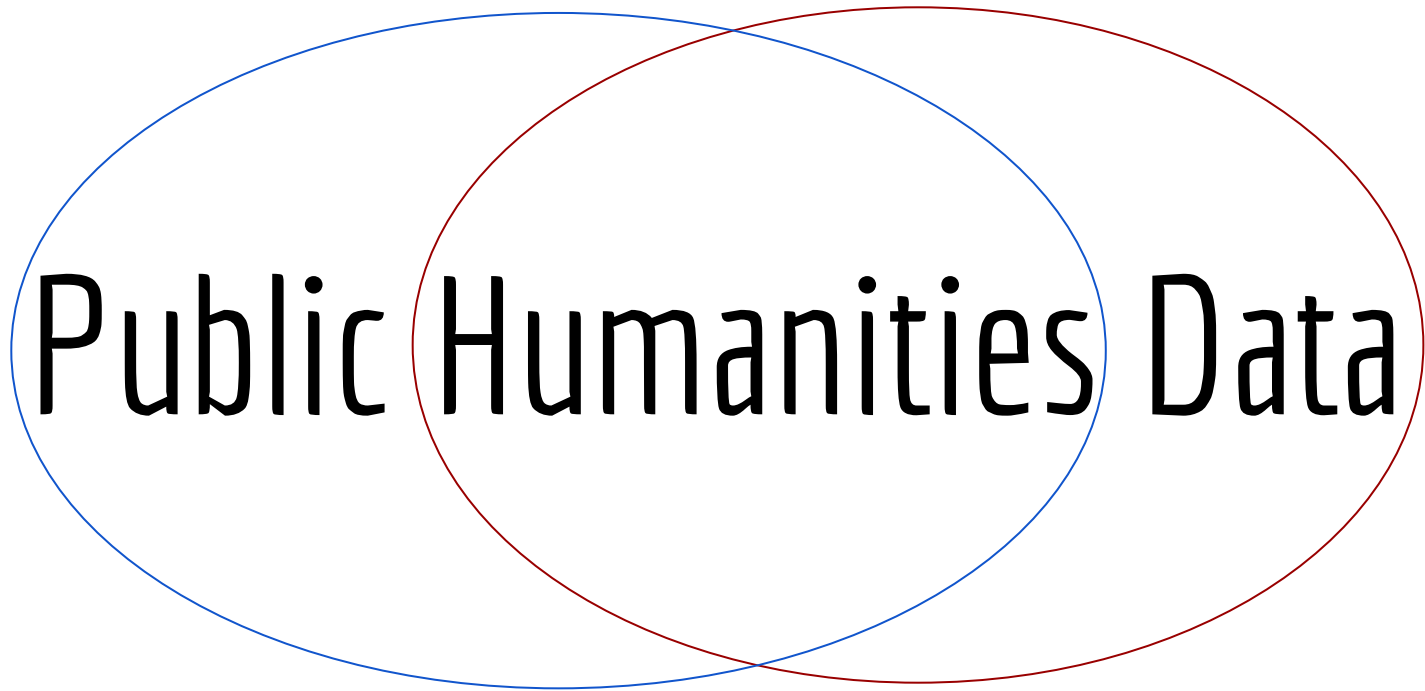
## Early English Books Online Text Creation Partnership



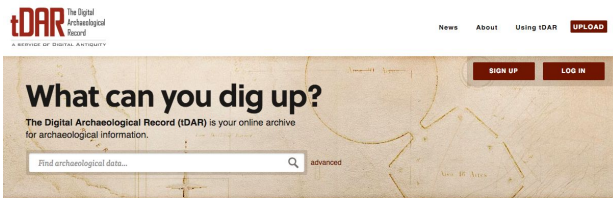
- Basic Search
- Boolean Search
- Proximity Search
- Bibliographic Search
- Word Index (Phase I)
- Word Index (Phase II)
- Browse (Phase I)
- Browse (Phase II)

About EEBO-TCP  
Demo EEBO-TCP  
comments to: [umdl@lib.umd.edu](mailto:umdl@lib.umd.edu)

[quod.lib.umich.edu/e/eebogroup](http://quod.lib.umich.edu/e/eebogroup)



# Group exercise!



## Access & Use

Broadening the access to archaeological data through simple search and browse functionality

LEARN MORE



## Preservation

Dedicated to ensuring long-term preservation of digital archaeological data.

LEARN MORE



## Upload Resources

Contribute documents, data sets, images, and other critical archaeological materials.

LEARN MORE



## Who Uses tDAR

Researchers like you. Uncover knowledge of the past, and preserve and protect archaeological resources.

LEARN MORE



This map shows murals from USC's Helen Drake Collection, UCLA's Nancy Tenney collection, and the Los Angeles Mural Conservancy

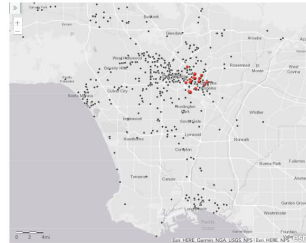
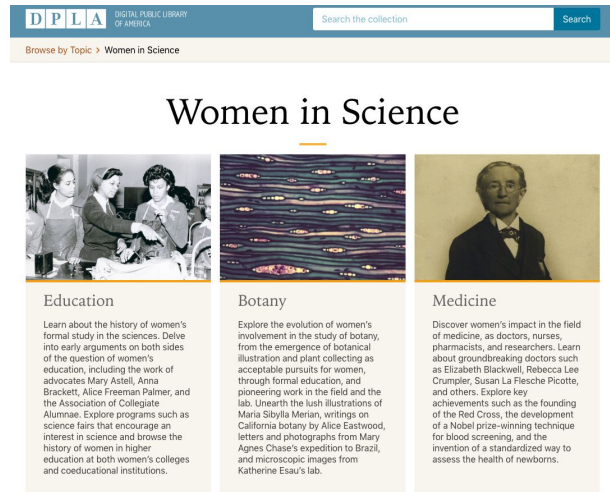


Image source: Andy Rothwell, taken at Vermont Ave. and W. 4th St. Los Angeles, CA. Map by Hendrik and created by Murals of LA. <https://muralcartographies.github.io/MuralsOfLA/>

<https://muralcartographies.github.io/MuralsOfLA/>



## Education

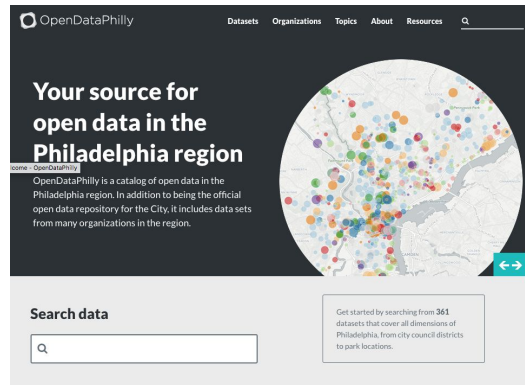
Learn about the history of women's formal study in the sciences. Delve into early arguments on both sides of the question of women's education, including the work of advocates Mary Astell, Anna Brackett, Alice Freeman Palmer, and the Association of Collegiate Alumnae. Explore programs such as science fairs that encourage an interest in science and browse the history of women in higher education at both women's colleges and coeducational institutions.

## Botany

Explore the evolution of women's involvement in the study of botany, from the emergence of botanical illustration and plant collecting as acceptable pursuits for women, through formal education, and pioneering work in the field and the lab. Unearth the lush illustrations of Maria Sibylla Merian, writings on California botany by Alice Eastwood, letters and photographs from Mary Agnes Chase's expedition to Brazil, and microscopic images from Katherine Esau's lab.

## Medicine

Discover women's impact in the field of medicine, as doctors, nurses, pharmacists, and researchers. Learn about groundbreaking doctors such as Elizabeth Blackwell, Rebecca Lee Crumpler, Susan La Flesche Picotte, and others. Explore key achievements such as the founding of the Red Cross, the development of a Nobel prize-winning technique for blood screening, and the invention of a standardized way to assess the health of newborns.

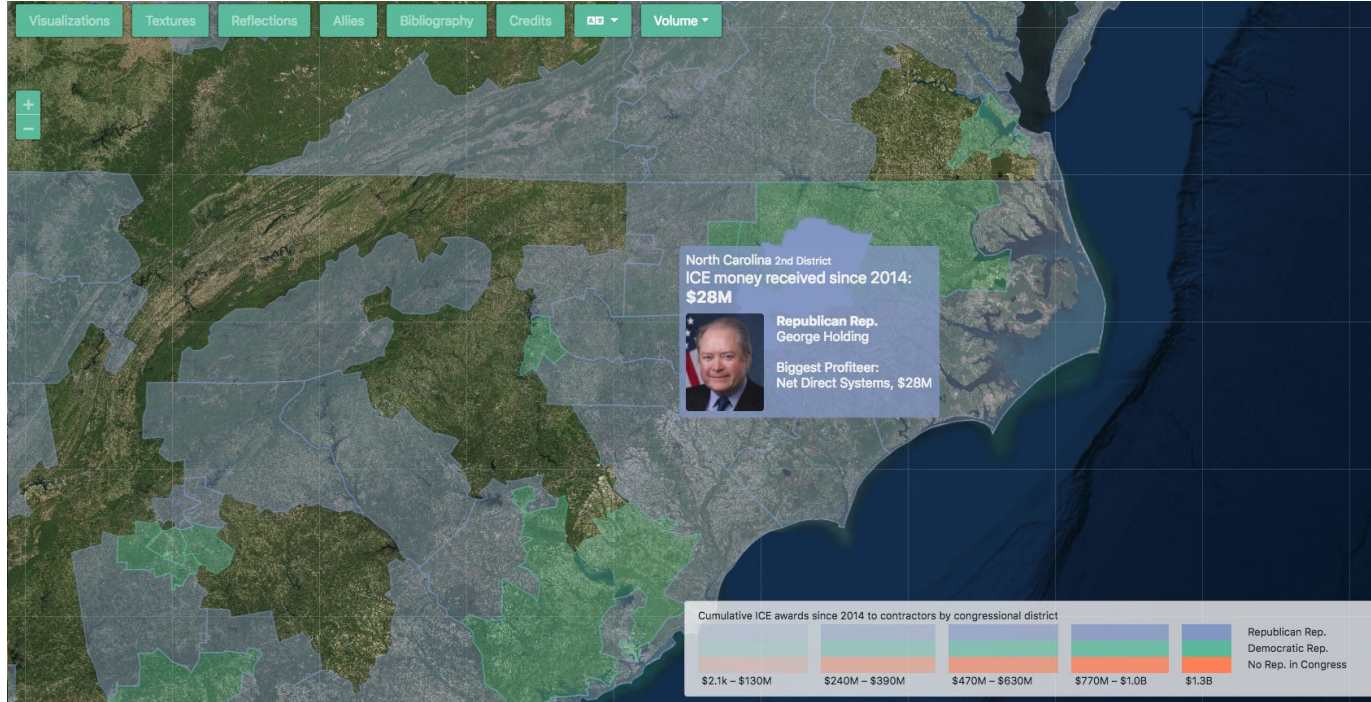


## Search data



Get started by searching from 361 datasets that cover all dimensions of Philadelphia, from city council districts to park locations.

# What are public humanities data for?



<http://xpmethod.plaintext.in/torn-apart/volume/2/index>





Data  $\approx$  STEM

Analysis  $\approx$  Interpretation

Public  $\approx$  Open



# Public Humanities + Humanities Data

Open (process, product)

For/Of/About Communities

Technolog(ies)

Continuum of expertise






# How do I do public humanities data?

Beg  
Borrow  
Steal

Gather  
Build  
Share



[< Back to Data & Visualization](#)
[< Back to Researcher Support](#)
**NEED HELP?**
[Request a  
Consultation](#) 


Alison Blaine



Erica Hayes



# Text and Data Mining

Text and data mining are the computer-based processes of extracting relevant information and/or patterns from machine-readable text or data. The work usually involves examining large data sets. Text and data mining is used across disciplines, from history to biology, computer science to political science.

## • Explore Data Sources

- [Scholarly APIs and Datasets](#)
- [Tools and Tutorials](#)

Depending on your discipline and the kind of data you want to mine, the process may be called something other than text mining or data mining. Related terms include: text data mining, text analytics, content mining, audio mining, software mining, image mining, metadata mining, and video mining.



# Ask Us

# lib.ncsu.edu/do/text-and-data-mining

## UPCOMING WORKSHOPS

SEP  
26

Introduction to  
APIs and JSON  
10:00 am to 11:30 am

SEP  
28

Simple Text  
Analysis with  
Voyant Tools and  
AntConc  
1:00 pm to 3:00 pm

#### NEED HELP?

Email Us →



Shaun Bennett



Will Cross



Erica Hayes



Micah  
Vandegrift



Mira Waller

# Digital Scholarship

Digital Scholarship extends traditional methods of research by leveraging technologies (such as visualization, GIS, text and data mining) to advance research, collaboration, teaching, and learning across disciplines. The NCSU Libraries can work with you to develop open techniques and creative practices that best support your research.

## Featured Digital Scholarship Projects at the NCSU Libraries:

- Immersive Scholar
- bookBot Poet
- Tracking Trethewy's Terror
- Chobson and the Dragons
- The North Carolina Architects & Builders: A Biographical Dictionary
- Shooting Wars: A Documentary of American Military Conflicts
- The Inside Wood Database
- North Carolina Geospatial Data Archiving Project
- Wolf Walk

## Get involved:

- Participate in our Open Project Incubator
- Attend our [Digital Scholarship Workshops](#).
- Re-use our [Workshop materials](#).

# Ask Us

## UPCOMING WORKSHOPS

SEP

5

What is Digital Scholarship?: An Introduction to Digital Research & Methods  
2:00 pm to 3:00 pm

SEP

7

BookBot Poet: Digital Research & Methods  
1:00 pm to 2:00 pm

SEP

14

Understand and Build Your Scholarly Identity  
10:00 am to 11:30 am

SEP

14

Introduction to XML and Text Encoding  
1:00 pm to 3:00 pm

[View Digital Scholarship Workshops →](#)

[lib.ncsu.edu/do/digital-scholarship](http://lib.ncsu.edu/do/digital-scholarship)

...extends traditional methods of research by leveraging technologies (such as visualization, GIS, text and data mining) to advance research, collaboration, teaching, and learning across disciplines.

digital scholarship @ NCSU Libraries

# Digital Scholarship Workshop Series

SEP

14

Introduction to XML and Text Encoding

September 14

1:00 PM to 3:00 PM

AT THE

ITTC Lab 2, D. H. Hill Library

OCT

24

Storytelling with GIS

October 24

1:00 PM to 3:00 PM

AT THE

ITTC Lab 2, D. H. Hill Library

OCT

26

Getting Started with Building Digital Exhibits in Omeka

October 26

10:00 AM to 12:00 PM

AT THE

ITTC Lab 2, D. H. Hill Library

[go.ncsu.edu/dsworkshops](http://go.ncsu.edu/dsworkshops)

---