Humanities Datasets in Context

Literature

African American Literature (1853-1923)

From: Melanie Walsh's list of Datasets

Data location	Format	Original Source	Notes
African American Literature (1853-1923)	Zip of txt files	Amardeep Singh	*Plain text of over 100 novels by Black literary authors including poetry and novels. Spreadsheet of metadata for the works.*

Additional information

Please follow the principle of the Colored Conventions Project when using and downloading the data

Colonial South Asian Literature (1850-1923)

From: Melanie Walsh's list of Datasets

Data location	Format	Original Source	Notes
Colonial South Asian Literature (1850-1923)	Zip of txt files	Amardeep Singh	*Plain text files of British and South Asian writers in English or in translation.*

TxtLab's Multilingual Novels (1771-1932)

From: Melanie Walsh's list of Datasets

Data location	Format	Original Source	Notes
Multilingual Novels Database	Zip of txt files	Andrew Piper, TxtLab	*Plain text files of 450 novels in German, French, and English*

Goodreads Ratings

From: Kaggle

Data location	Format	Original Source	Notes
Goodreads data	CSV	Goodreads API and Kaggle user Soumik	*CSV containing metadata about the book and the average rating on Goodreads*

Library Data

Shakespeare & Company: Lending Library Data

CDH

Data location	Format	Original Source	Notes

| **Shakespeare & Co. Lending Library Data** | CSV and JSON files | The Shakespeare and Company Project | *CSV files of lending library borrows, purchases. memberships and renewals* |

Seattle Public Library Ciculation Data (2005-present)

Data location	Format	Original Source	Notes
Seattle Library checkouts by title 2005-present	exportable spreadsheet	City of Seattle	*Data for all checkouts since 2005, filterable by interest*

Pop Culture

Game of Thrones Character Relationships

Data location	Format	Original Source	Notes
GoT Character data	CSV files	A. Beveridge and J. Shan	*107 characters and their interactions. Best suited for Network Analysis.*

Bechdel Test for Movies

Data location	Format	Original Source	Notes
Bechdel Movie Data	CSV	"The Dollars-and-Cents Case Against Hollywood's Exclusion of Women"	*1,615 films from 1990 to 2013 measured against the Bechdel Test.*

Hollywood Film Dialogue by Character Gender and Age

Data location	Format	Original Source	Notes
Hollywood Film Dialogue	CSV	"Film Dialogue from 2,000 screenplays, Broken Down by Gender and Age".	*Dialogue for 2000 films (1925-2015). For more info see FAQ for the "Film Dialogue, By Gender" Project.*

50 Years of Pop Music Lyrics

Data location	Format	Original Source	Notes
Billboard Year-End Hot 100 (1965-2015)	CSV	Kaylin Pavlik	*50 years of lyrics from Billboard's Top 100. Includes song, artist, year, lyrics, and source.*

Graphic Novel Corpus

Data location	Format	Original Source	Notes
Graphic Novel Corpus	Zip of CSV files and PDF of database logic	About the Graphic Novel Corpus	*240 graphic novels. Files include author, title, and other metadata. Structured in multiple CSV files for creation of a database. Charts section shows possible analysis.*

Miscellaneous

Queer Politics at Princeton

From: QP@P

Data location	Format	Original Source	Notes
Queer Politics at Princeton	CSV	Andrew Reynolds, Queer Politicians Data, QP@P Princeton	*Dataset includes out LGBTQI+ elected officials since 1976, geographic information (w/o Lat/Long), political party, role, etc.*

The Endangered Languages Project

From: Endangered Languages Project

Data location	Format	Original Source	Notes
Endangered Languages Data	CSV	Alliance for Linguistic Diversity	*Dataset includes vitality of a language, linguistic details, number of living speakers and geographic data.*

Who Has Your Face?

From: Atlas of Surveillance

Data location	Format	Original Source	Notes
Facial Recognition Data	CSV	Electric Frontier Foundation Who Has Your Face? report	*Dataset organized by state that details which agencies have access to which type of facial recognition data*

Torn Apart/Separados

From: Mobilized Humanities

Data location	Format	Original Source	Notes
Open Data Repository	CSV files	Torn Apart/Separados	*Data culled from multiple sources to compile an atlas of ICE detention centers on the southern border*

Nobel Prize Winners

From: Melanie Walsh's list of Datasets

Data location	Format	Original Source	Notes
Nobel Prize Winners data	CSV	The European Data Portal and the official Nobel Prize API	*957 Nobel Prize winners (1901-2017) includes metadata about the winners and location data for mapping*

What's on the Menu?

From: Melanie Walsh's list of Datasets

Data location	Format	Original Source	Notes
What's on the Menu data	CSV	New York Public Library	*Menu and dish data including prices collected so far from the 45,000 menus in their colllection (1840-present)*

Additional information

This project was the subject of Katie Rawson and Trevor Muñoz' article "Against Cleaning"