



Open Data Boot Camp

SESSION 2

Wende A. Mix Ph.D.

Session 2: Data Resources for Research (February 9, 2022)



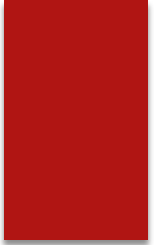
FINDING OPEN DATA SOURCES



ASSESSING APPROPRIATENESS
AND QUALITY OF DATA

Process - Data Quality



- 
- If there is a README or description page for the dataset, what information does it provide?
 - How is the information in the dataset organized?
 - What relationships might there be between different parts of the dataset?
 - What could the dataset be used for?

From Teaching with the Library of Congress Blog

Questions About Open Data Sites

- ▶ Is there metadata (possibly more detailed than a readme page)?
 - ▶ Who “owns” the data?
 - ▶ Are all fields (columns) defined?
- ▶ Is the data authoritative?
- ▶ When was the data last updated?
- ▶ Can you download the data?
 - ▶ What formats (Excel, csv, JSON, GeoJSON, shp, Odata....)
 - ▶ Are you able to download all relevant records or a partial set?
- ▶ Can you filter the data?
- ▶ Can you create maps and graphs on the site using the data?
 - ▶ If the data are presented as maps or graphs, can you access the raw data?

Open Data Resources



[Teaching with the Library of Congress](#)

[Datasets at the Library of Congress](#)

[Google search - example poverty open data](#)

[ICPSR - UMICH](#)

[Socrata/ Open Data Network](#)

Some Open Data Sources

- ▶ Local/city government: [Washington DC](#)
- ▶ State government: [New York](#)
- ▶ Federal agencies – [data.gov](#), [NHGIS/IPUMS](#), [internet archive.org](#),
- ▶ Private business/ non-profits datasets ([AirBNB](#), [YELP](#), [Kaggle](#),
[WorldBank](#), [World Health Organization](#), [Google Public Data Explorer](#),
[Registry of Open Data on AWS](#))