



What Is “**Data**” in the **Humanities**?

October 6, 2021

Sierra Eckert
Matthew Chandler

The Humanities
Data Workshop
Series



What we'll learn today

1. Data: Some Humanist Definitions
2. From Sources to Data
3. Anatomy of a Humanities Dataset: Data? Metadata? Collection? Archive?:
4. What does a humanities approach to data look like?
5. Q&A (Ask Us Anything)

1.

“Data”: Some Humanist Definitions

Etymologies

- **data** | *datum* — to give | rhetorical

Etymologies

- **data** | *datum* — to give | rhetorical
- **fact** | *facere* — to do | ontological

Etymologies

- **data** | *datum* — to give | rhetorical
- **fact** | *facere* — to do | ontological
- **evidence** | *videre* — to see | epistemological

- **data** | *datum* — to give | rhetorical
- **fact** | *facere* — to do | ontological
- **evidence** | *videre* — to see | epistemological

“...the meaning of data must always shift with argumentative strategy and context—and with the history of both.”

Daniel Rosenberg, “Data Before the Fact,” in *“Raw Data” Is an Oxymoron*, ed. Lisa Gitelman (MIT Press, 2013)

Google Books Ngram Viewer

Q data,evidence,fact



1700 - 2019 ▾

English (2019) ▾

Case-Insensitive

Smoothing ▾

 GALE PRIMARY SOURCES

Eighteenth Century Collections Online

Search...




Authors


About


Research Tools

SEARCH OPTIONS

Advanced Search

[Topic Finder](#)

Tools used (and critically examined) in Daniel
Rosenberg's essay "Data Before the Fact," (2013)

Let's do a quick exercise:

“DATA” and GOOGLE
NGRAMS

- Go to <https://books.google.com/ngrams>
- Type in Rosenberg's search terms:
“data,” “evidence,” “fact”

Google Books Ngram Viewer

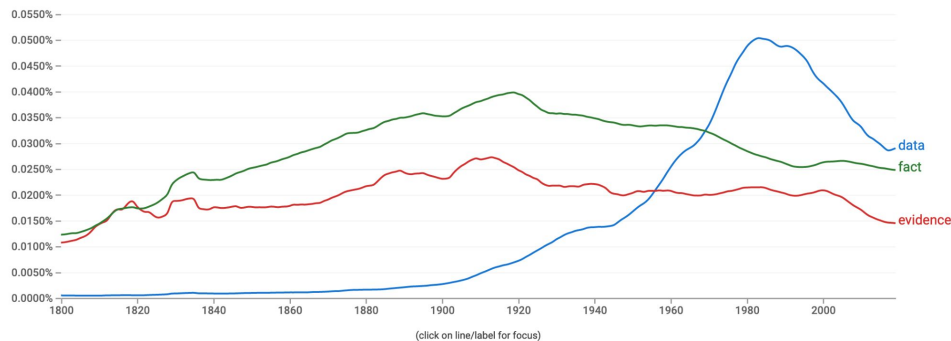
data,evidence,fact

1800 - 2019

English (2019)

Case-Insensitive

Smoothing



Search in Google Books

data

>

1800 - 1930

1931 - 1987

1988 - 1993

1994 - 2008

2009 - 2019

English (2019)

- What data does the GoogleBooks NGram Viewer rely on?
- What else would we have to know before using this tool to make historical arguments?

- **data** | *datum* — to give
- **capta** | *capiō* — to take

- **data** | *datum* — to give
- **capta** | *capiō* — to take

“*Capta* is ‘taken’ actively while *data* is assumed to be a ‘given’ able to be recorded and observed. From this distinction, a world of differences arises. Humanistic inquiry acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, taken, not simply given as a natural representation of pre-existing fact.”

Johanna Drucker, “Humanities Approaches to Graphical Display,” *Digital Humanities Quarterly* 5, no. 1 (March 10, 2011).

What kind of things make for meaningful data?

Your data is only as good as the categories you capture

Silent Film & Melodramatic Conventions

Title	Year	Virtuous heroine	Cruel villain	Terrible evil	Heroine in peril	Broad gestures
The Lonedale Operator	1911	x	x	x	x	x
Birth of a Nation	1915	x	x	x	x	x
Get Rich Quick	1912	x				x

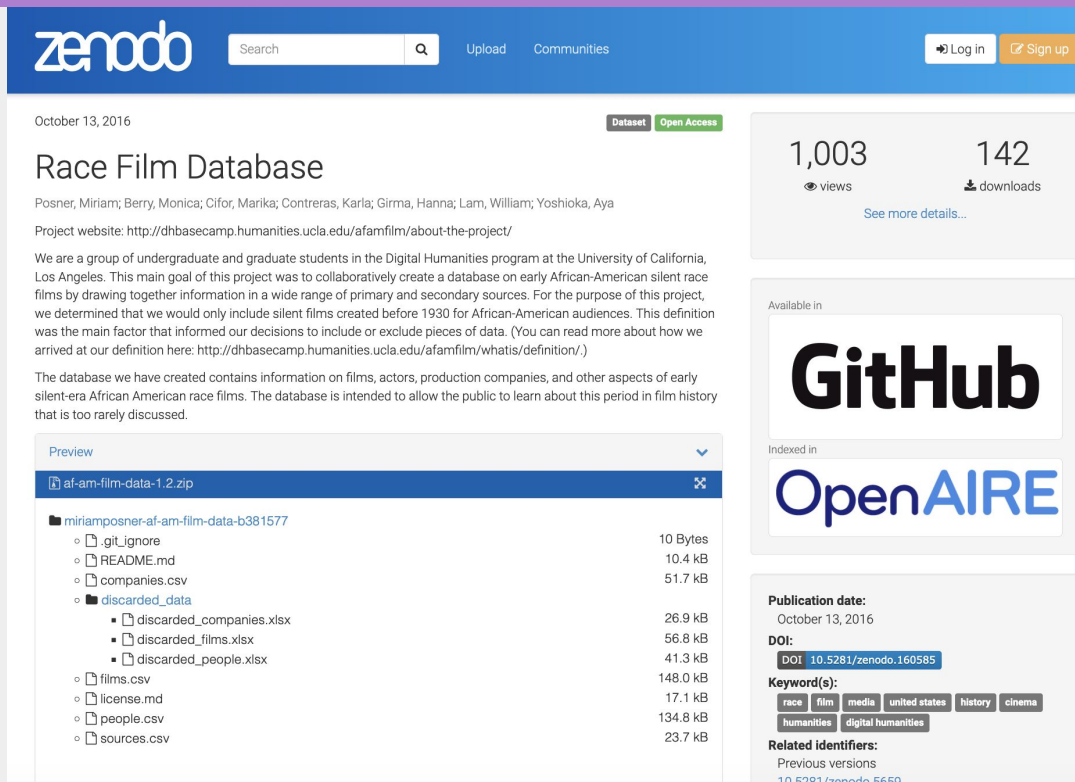
Miriam Posner's hypothetical silent film dataset

Your data is only as good as the categories you capture

Miriam Posner's team's actual dataset

Sources used:

- ❖ UCLA's George P. Johnson Negro Film Collection archival records
 - Newspaper and magazine reviews
 - Film and video records
- ❖ Published filmographies, like *African American Films Through 1959: A Comprehensive, Illustrated Filmography* (1988)



The screenshot shows the Zenodo dataset page for 'Race Film Database'. The page header includes the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. The dataset is dated October 13, 2016, and has 1,003 views and 142 downloads. The dataset is titled 'Race Film Database' and is described as a project website for the Digital Humanities program at the University of California, Los Angeles. The description states that the project aims to collaboratively create a database of early African-American silent race films by drawing together information from a wide range of primary and secondary sources. The dataset is available in a preview mode, showing a file named 'af-am-film-data-1.2.zip' with a size of 10.4 kB. The file is a zip archive containing a folder named 'miriamposner-af-am-film-data-b381577' which includes files like '.gitignore', 'README.md', 'companies.csv', 'discarded_data', 'discarded_companies.xlsx', 'discarded_films.xlsx', 'discarded_people.xlsx', 'films.csv', 'license.md', 'people.csv', and 'sources.csv'. The dataset is indexed in GitHub and OpenAIRE. The publication date is October 13, 2016, and the DOI is 10.5281/zenodo.160585. The keyword(s) are race, film, media, united states, history, cinema, humanities, and digital humanities. Related identifiers include previous versions 10.5281/zenodo.5659.

zenodo

Search

Upload Communities

Log in Sign up

October 13, 2016

Dataset Open Access

Race Film Database

Posner, Miriam; Berry, Monica; Cifor, Marika; Contreras, Karla; Girma, Hanna; Lam, William; Yoshioka, Aya

Project website: <http://dhbasecamp.humanities.ucla.edu/afamfilm/about-the-project/>

We are a group of undergraduate and graduate students in the Digital Humanities program at the University of California, Los Angeles. This main goal of this project was to collaboratively create a database on early African-American silent race films by drawing together information in a wide range of primary and secondary sources. For the purpose of this project, we determined that we would only include silent films created before 1930 for African-American audiences. This definition was the main factor that informed our decisions to include or exclude pieces of data. (You can read more about how we arrived at our definition here: <http://dhbasecamp.humanities.ucla.edu/afamfilm/whatis/definition/>.)

The database we have created contains information on films, actors, production companies, and other aspects of early silent-era African American race films. The database is intended to allow the public to learn about this period in film history that is too rarely discussed.

Preview

af-am-film-data-1.2.zip

miriamposner-af-am-film-data-b381577

- .gitignore 10 Bytes
- README.md 10.4 kB
- companies.csv 51.7 kB
- discarded_data
 - discarded_companies.xlsx 26.9 kB
 - discarded_films.xlsx 56.8 kB
 - discarded_people.xlsx 41.3 kB
- films.csv 148.0 kB
- license.md 17.1 kB
- people.csv 134.8 kB
- sources.csv 23.7 kB

Available in

GitHub

Indexed in

OpenAIRE

Publication date:

October 13, 2016

DOI:

DOI 10.5281/zenodo.160585

Keyword(s):

race film media united states history cinema humanities digital humanities

Related identifiers:

Previous versions

10.5281/zenodo.5659

Your data is only as good as the categories you capture

Categories of data in Miriam Posner's team's dataset:

Title
Genre
Primary Source Verification
Production Company
Production date
Distribution date
Length (reels)
Actors
Directors
Producers
Camerawork
Source
Writer
Otherwise Involved
Link to media
Distributor

The screenshot shows the Zenodo dataset page for 'Race Film Database'. The header includes the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. The dataset is dated October 13, 2016, and is marked as 'Dataset' and 'Open Access'. It has 1,003 views and 142 downloads. The project website is listed as <http://dhbasecamp.humanities.ucla.edu/afamfilm/about-the-project/>. The description states that the project is a collaborative effort by undergraduate and graduate students at the University of California, Los Angeles, to create a database of early African-American silent race films. The database includes information on films, actors, production companies, and other aspects of early silent-era African American race films. A file preview section shows a folder named 'miriamposner-af-am-film-data-b381577' containing various files like '.git_ignore', 'README.md', 'companies.csv', 'discarded_data' (a subfolder with 'discarded_companies.xlsx', 'discarded_films.xlsx', and 'discarded_people.xlsx'), 'films.csv', 'license.md', 'people.csv', and 'sources.csv'. The dataset is available in GitHub and indexed in OpenAIRE. The publication date is October 13, 2016, and the DOI is 10.5281/zenodo.160585. Keywords include race, film, media, united states, history, cinema, humanities, and digital humanities. Related identifiers include previous versions 10.5281/zenodo.5559.

zenodo

Search

Upload Communities

Log in Sign up

October 13, 2016

Dataset Open Access

Race Film Database

Posner, Miriam; Berry, Monica; Cifor, Marika; Contreras, Karla; Girma, Hanna; Lam, William; Yoshioka, Aya

Project website: <http://dhbasecamp.humanities.ucla.edu/afamfilm/about-the-project/>

We are a group of undergraduate and graduate students in the Digital Humanities program at the University of California, Los Angeles. This main goal of this project was to collaboratively create a database on early African-American silent race films by drawing together information in a wide range of primary and secondary sources. For the purpose of this project, we determined that we would only include silent films created before 1930 for African-American audiences. This definition was the main factor that informed our decisions to include or exclude pieces of data. (You can read more about how we arrived at our definition here: <http://dhbasecamp.humanities.ucla.edu/afamfilm/whatis/definition/>.)

The database we have created contains information on films, actors, production companies, and other aspects of early silent-era African American race films. The database is intended to allow the public to learn about this period in film history that is too rarely discussed.

Preview

af-am-film-data-1.2.zip

- miriamposner-af-am-film-data-b381577
 - .git_ignore 10 Bytes
 - README.md 10.4 kB
 - companies.csv 51.7 kB
 - discarded_data
 - discarded_companies.xlsx 26.9 kB
 - discarded_films.xlsx 56.8 kB
 - discarded_people.xlsx 41.3 kB
 - films.csv 148.0 kB
 - license.md 17.1 kB
 - people.csv 134.8 kB
 - sources.csv 23.7 kB

Available in

GitHub

Indexed in

OpenAIRE

Publication date:
October 13, 2016

DOI:
DOI 10.5281/zenodo.160585

















Keyword(s):
race film media united states history cinema
humanities digital humanities

Related identifiers:
Previous versions
10.5281/zenodo.5559

2.

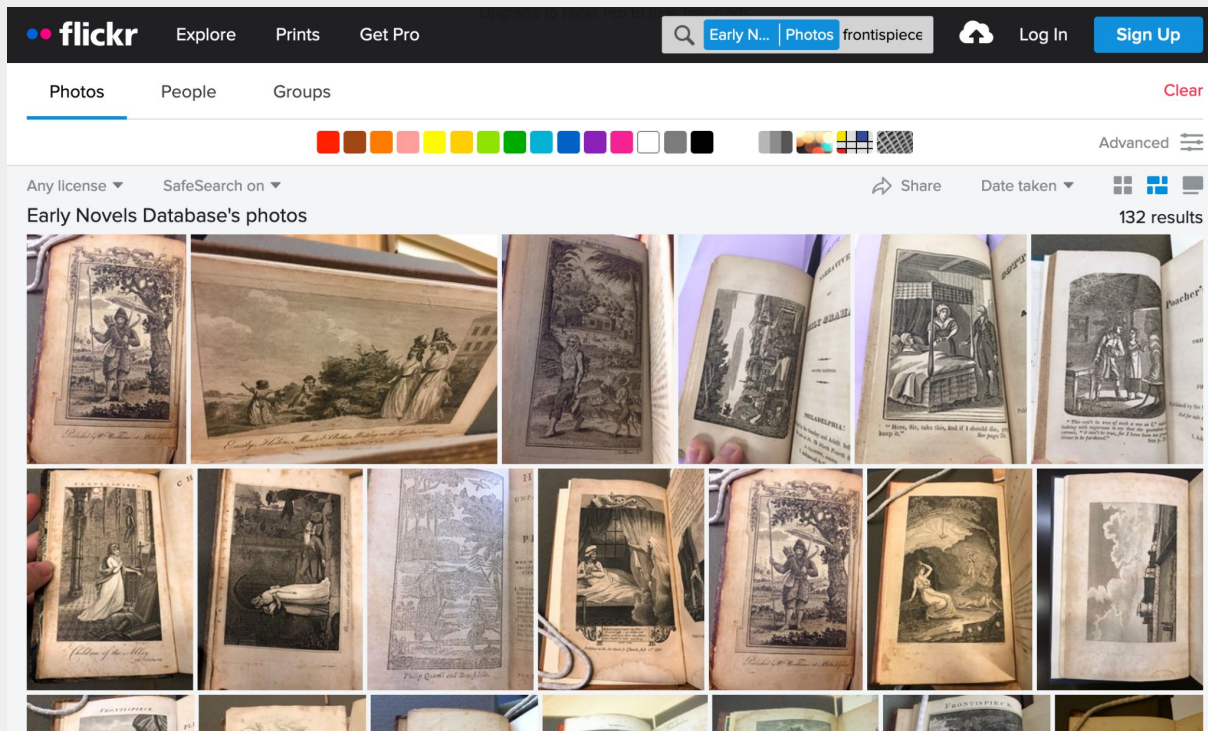
From Sources to Data

HUMANITIES DATA in the WILD

Name	
	2018-10-29-BL_2881.JPG
	2018-10-29-BL_2882.JPG
	2018-10-29-BL_2883.JPG
	2018-10-29-BL_2884.JPG
	2018-10-29-BL_2885.JPG
	2018-10-29-BL_2886.JPG
	2018-10-29-BL_2887.JPG
	2018-10-29-BL_2888.JPG
	2018-10-29-BL_2889.JPG
	2018-10-29-BL_2890.JPG
	2018-10-29-BL_2891.JPG
	2018-10-29-BL_2892.JPG
	2018-10-29-BL_2893.JPG
	2018-10-29-BL_2894.JPG
	2018-10-29-BL_2895.JPG
	2018-10-29-BL_2896.JPG


A sample archival dataset

HUMANITIES DATA in the WILD



Early Novels Database frontispiece dataset

HUMANITIES DATA in the WILD

 **CHRONICLING AMERICA**
Historic American Newspapers

Search America's historic newspaper pages from 1777-1963 or use the U.S. Newspaper Directory to find information about American newspapers published between 1690-present. Chronicling America is sponsored jointly by the National Endowment for the Humanities and the Library of Congress. [Learn more >](#)

Search Pages Advanced Search All Digitized Newspapers 1777-1963 **US Newspaper Directory, 1690-Present**

All states + from 1918 to 1963 + "Spanish flu" **GO**

Pages Available: 18,587,506 [Print](#) [Subscribe](#) [Share/Save](#) [Give Feedback](#)




Results 1 - 20 of 2779 1 | 2 | 3 | ... | 137 | 138 | 139 Jump to page: **GO** **View:** Gallery | List

2779 results containing **"Spanish flu"**

☐ Show only front pages Sort by: **Relevance** Results per page: **20**

The North Mississippi herald. (Water Valley, Yalobusha Co., Miss.), November 08, 1918, [Image 7](#)
Evening journal. [volume] (Wilmington, Del.), October 17, 1918, Page 8, [Image 8](#)
Omaha daily bee. (Omaha [Neb.]), October 09, 1918, [Image 5](#)
The Washington times. [volume] (Washington [D.C.]), October 29, 1918, FINAL EDITION, SPORTING PAGE, [Image 14](#)
The Bourbon news. [volume] (Paris, Ky.), October 22, 1918, Page PAGE TWO, [Image 2](#)
Courier Democrat. (Langdon, N.D.), October 30, 1919, [Image 8](#)
Audubon County journal. (Exira, Iowa), November 28, 1918, [Image 7](#)
Omaha daily bee. (Omaha [Neb.]), October 10, 1918, Page 10, [Image 10](#)
The Chickasha daily express. (Chickasha, Indian Territory [Okla.]), November 30, 1920, Page PAGE FIVE, [Image 5](#)
The Bismarck tribune. [volume] (Bismarck, N.D.), November 11, 1918, [Image 5](#)
Alexandria gazette. [volume] (Alexandria, D.C.), August 09, 1918, [Image 2](#)
Courier Democrat. (Langdon, N.D.), October 24, 1918, [Image 5](#)
Courier Democrat. (Langdon, N.D.), November 07, 1918, [Image 5](#)
Courier Democrat. (Langdon, N.D.), October 31, 1918, [Image 5](#)
The Mt. Sterling advocate. (Mt. Sterling, Ky.), November 05, 1918, [Image 8](#)

Sample full-text search of Chronicling America newspaper database




  

Info **Notes** **Tags** **Related**

Citation Key: Flu1918

Item Type Newspaper Article

Title The Flu

▼ Author (last) , (first)   

Abstract

Publication **Evening Journal**

Place Wilmington, Del

Edition

Date 1918/10/17 y m d

Section

Pages 8

Language eng

Short Title

ISSN 2641-6123

URL <https://chroniclingamerica.loc.gov/lccn/...>

Accessed 10/5/2021, 4:14:43 PM

Archive

Loc. in Archive

Library Catalog chroniclingamerica.loc.gov

Call Number

Rights Text is U.S. Government Work

Extra Issue: 1918/10/17
Publication: Evening journal. [volume]
Publisher: J. Milton Davidson

Zotero record for newspaper article

Exercise:

What are your sources?

Share a snapshot or sample of your source material on the GoogleDoc

GoogleDocs for break rooms:

Room 1: tinyurl.com/fvs5b22d

Room 2: tinyurl.com/4zzum4pn

Room 3: tinyurl.com/27575vm3

Room 4: tinyurl.com/jcfkp8cx

3. Anatomy of a Humanities Dataset:

Data? Metadata?
Collection? Archive?



Data

Metadata

Data

- ❖ Content
- ❖ Files representing artifacts or source materials (texts, artworks, music files)
- ❖ Examples:
 - Text files of articles or works of literature
 - Image files of archival objects or artworks
 - Oral history audio recordings or transcripts

Metadata

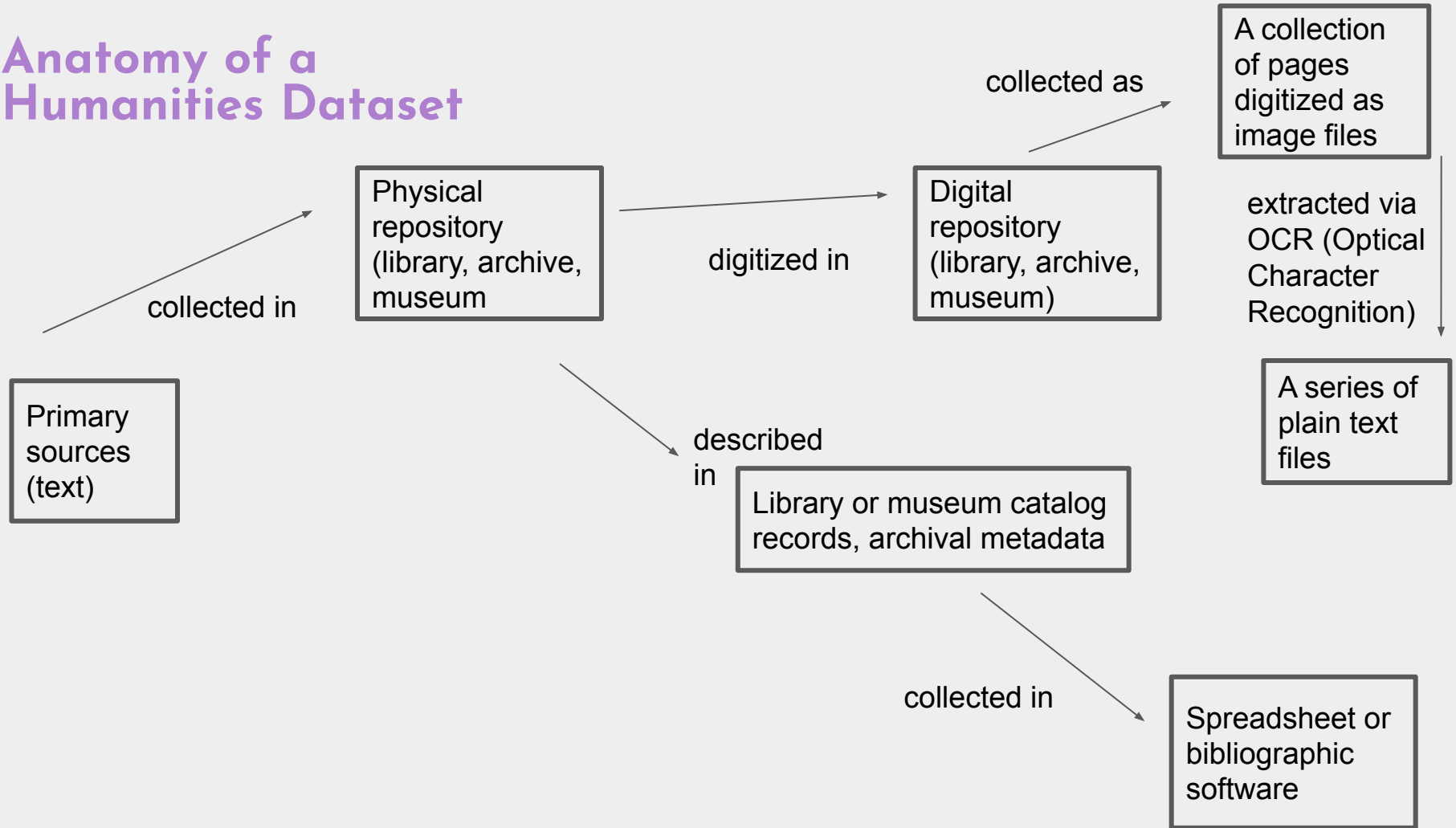
Data

- ❖ Content
- ❖ Files representing artifacts or source materials (texts, artworks, music files)
- ❖ Examples:
 - Text files of articles or works of literature
 - Image files of archival objects or artworks
 - Oral history audio recordings or transcripts

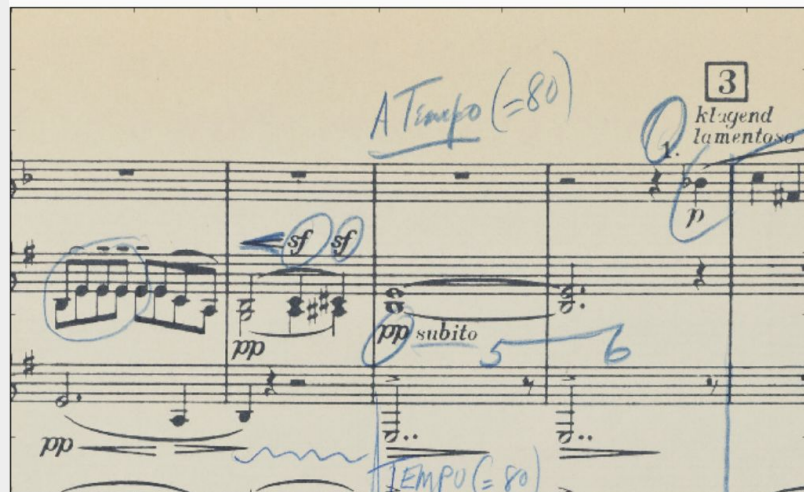
Metadata

- ❖ Context: “data about data”
- ❖ Files representing *contextual information about* artifacts or sources (titles, creators, date of creation, size, provenance)
- ❖ Generally machine-readable
- ❖ Examples:
 - Spreadsheets
 - Library catalog records
 - Encoded archival data

Anatomy of a Humanities Dataset



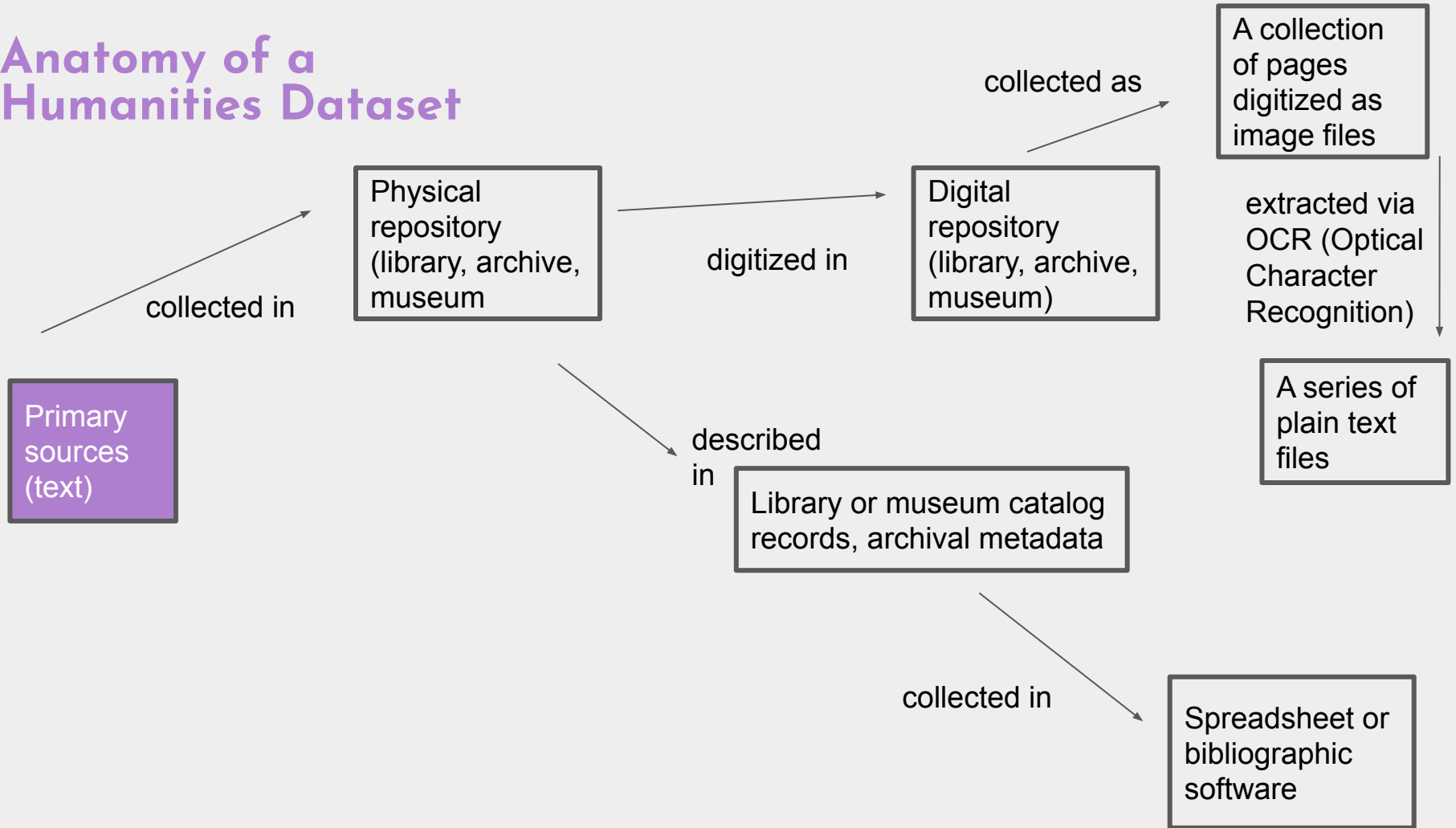
Example dataset:



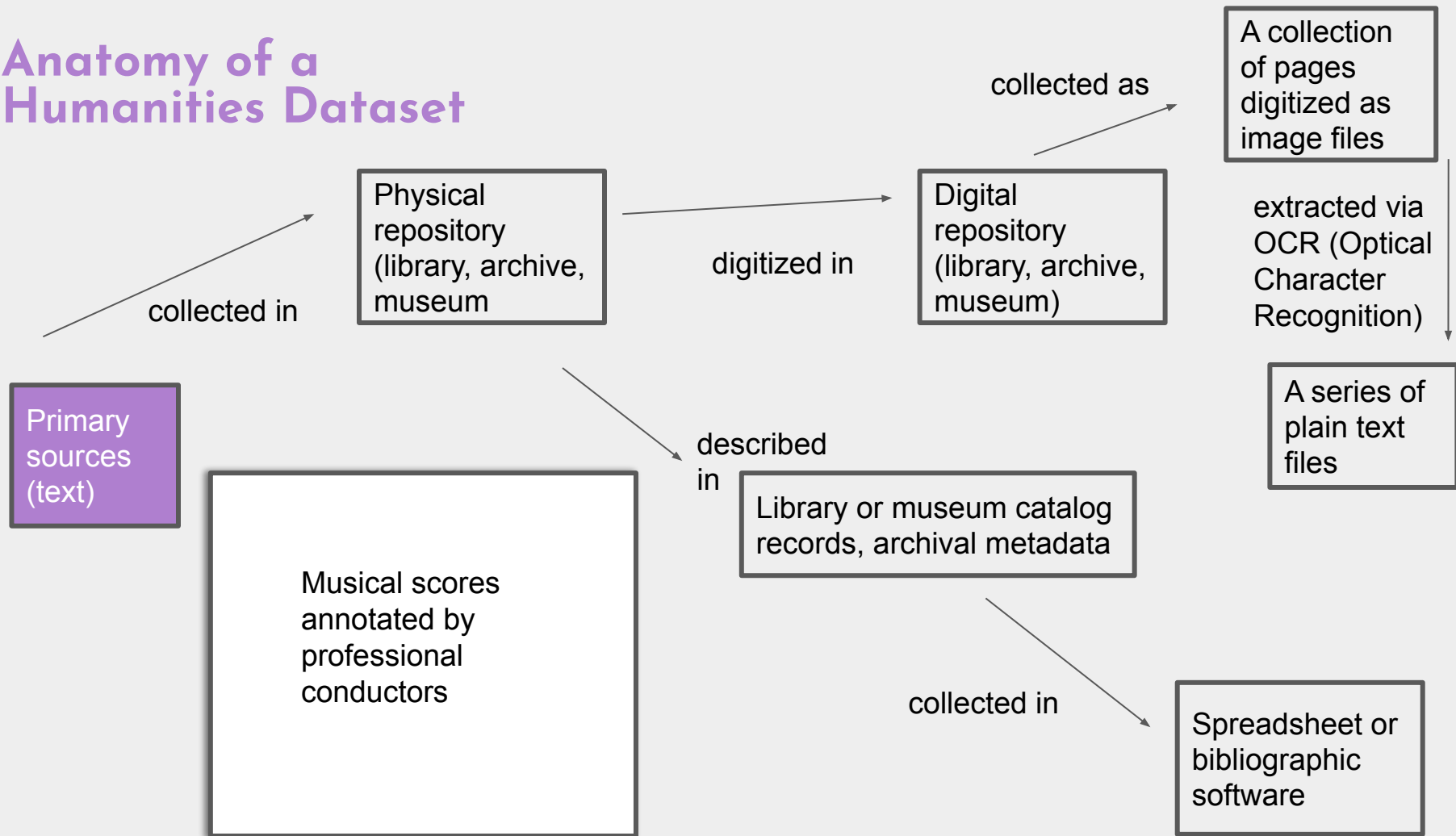
NY conductors' handwritten annotations on musical scores

Eamonn Bell and Laurent Pugin

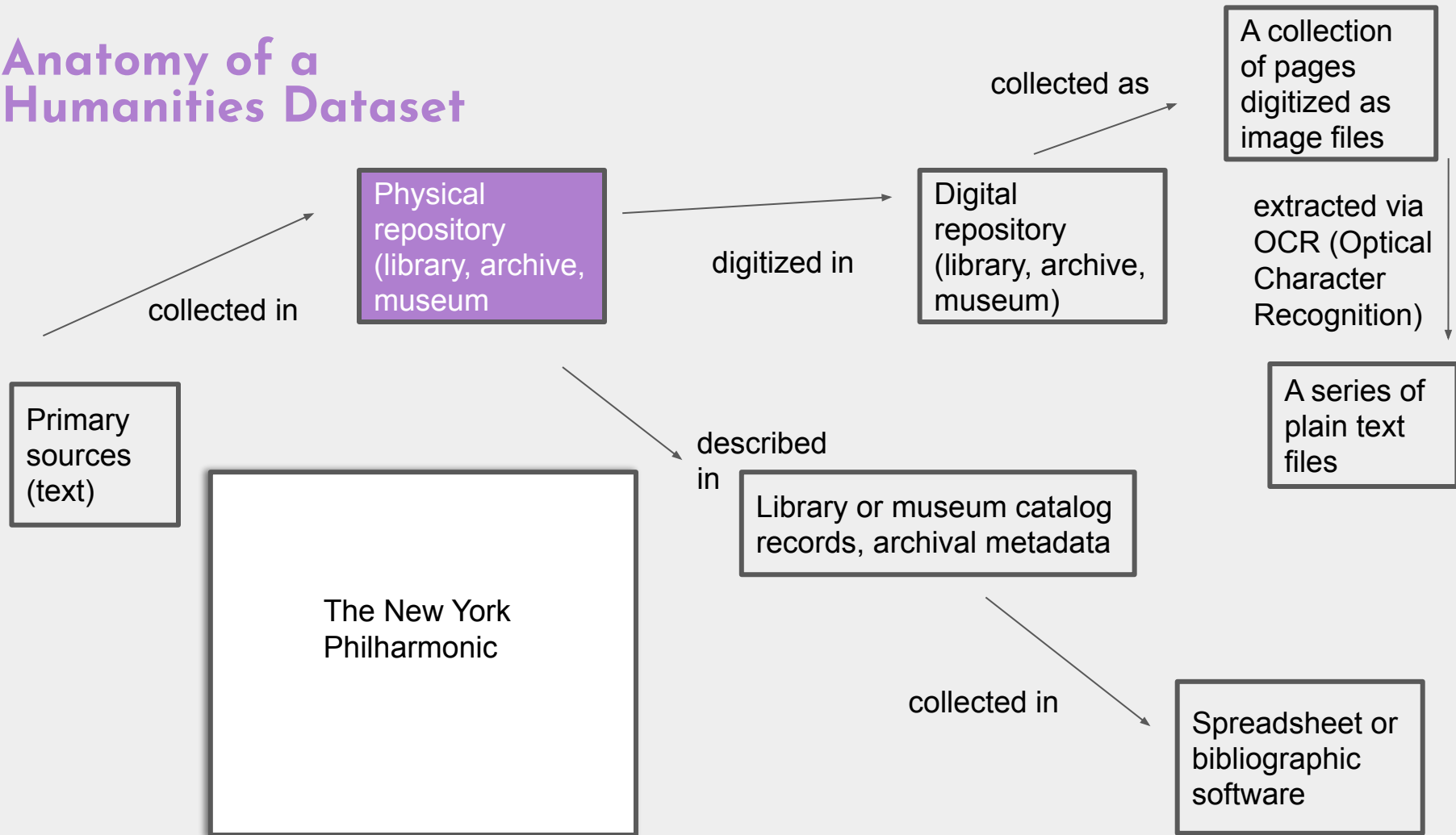
Anatomy of a Humanities Dataset



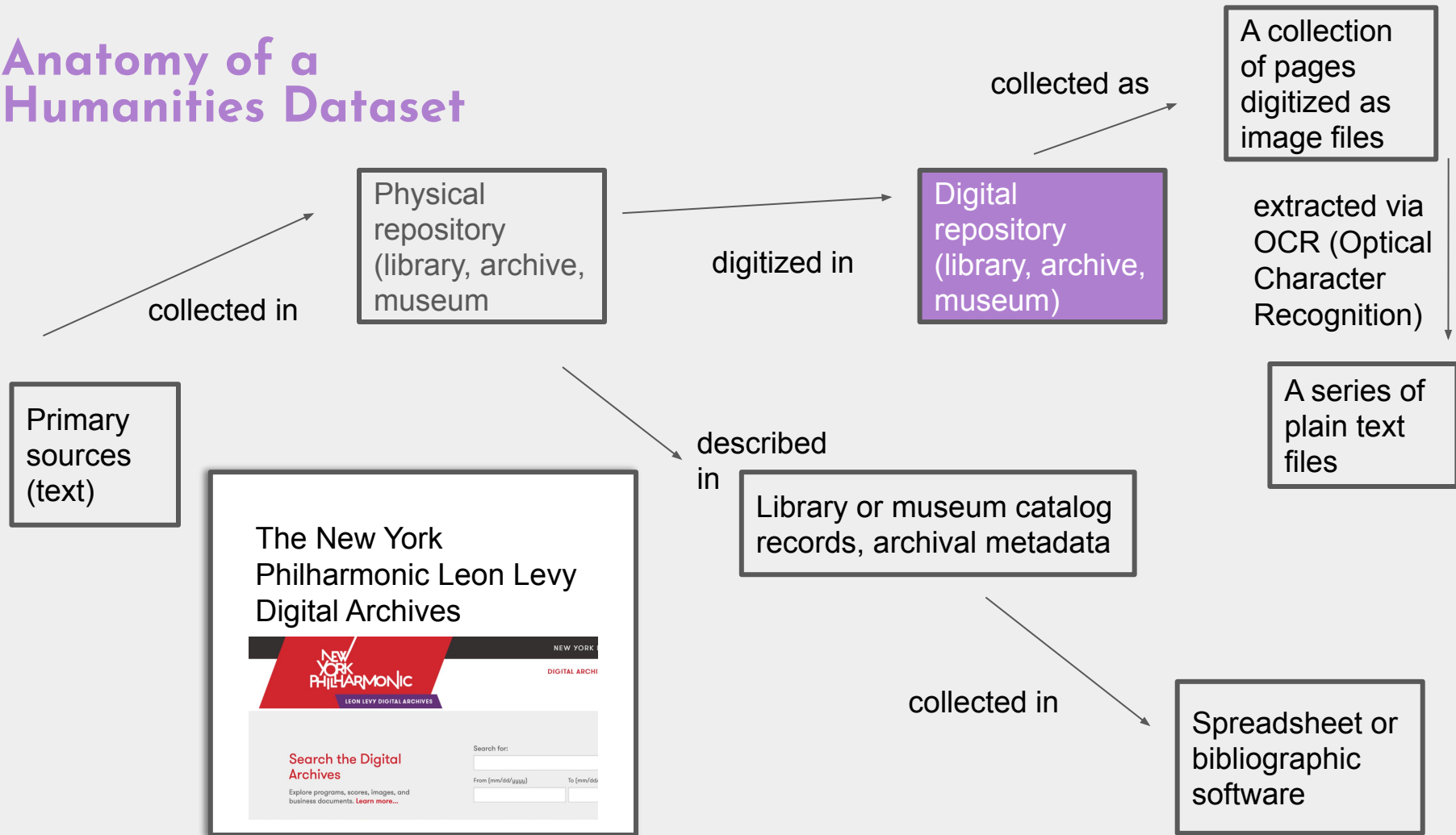
Anatomy of a Humanities Dataset



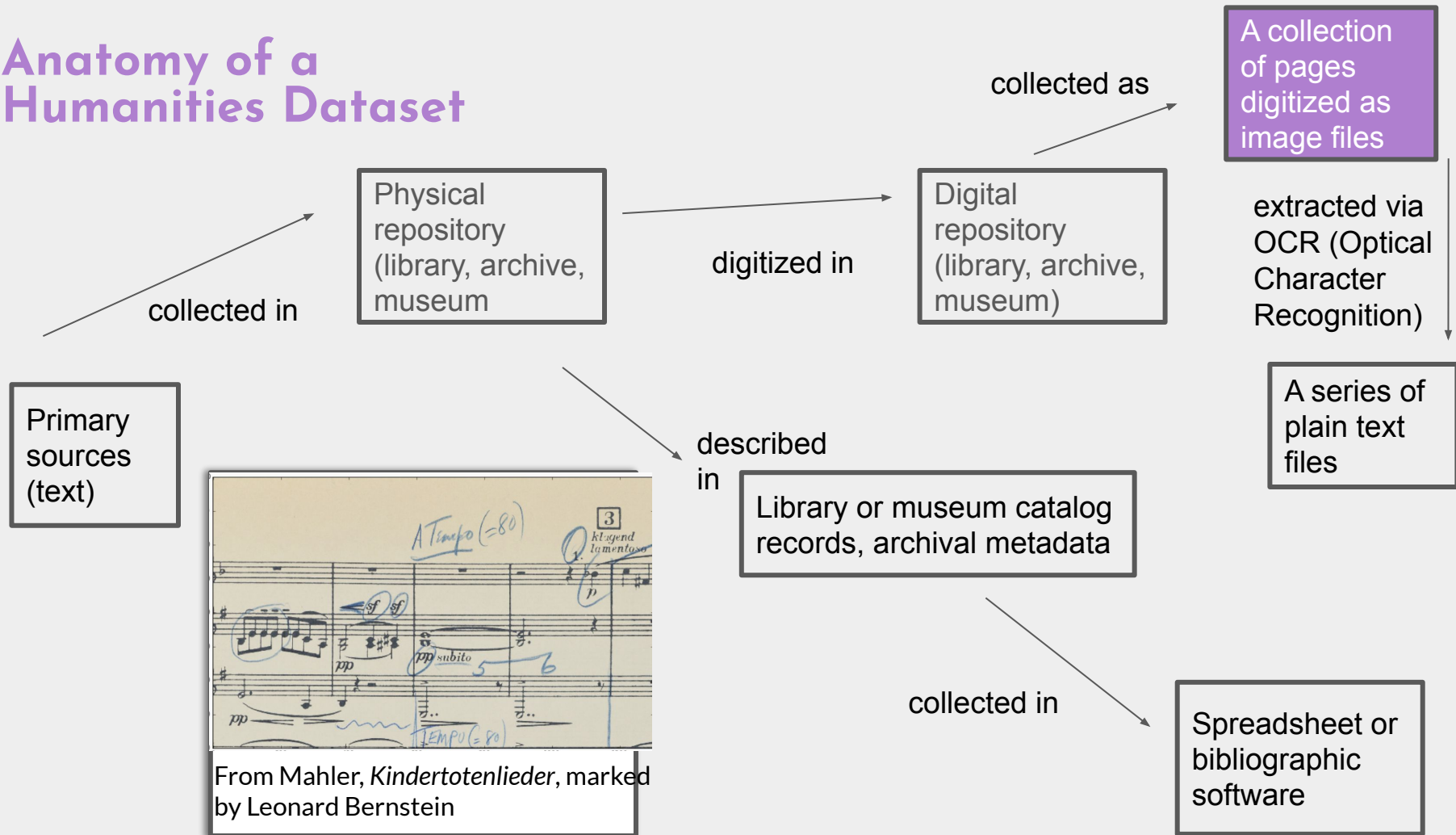
Anatomy of a Humanities Dataset



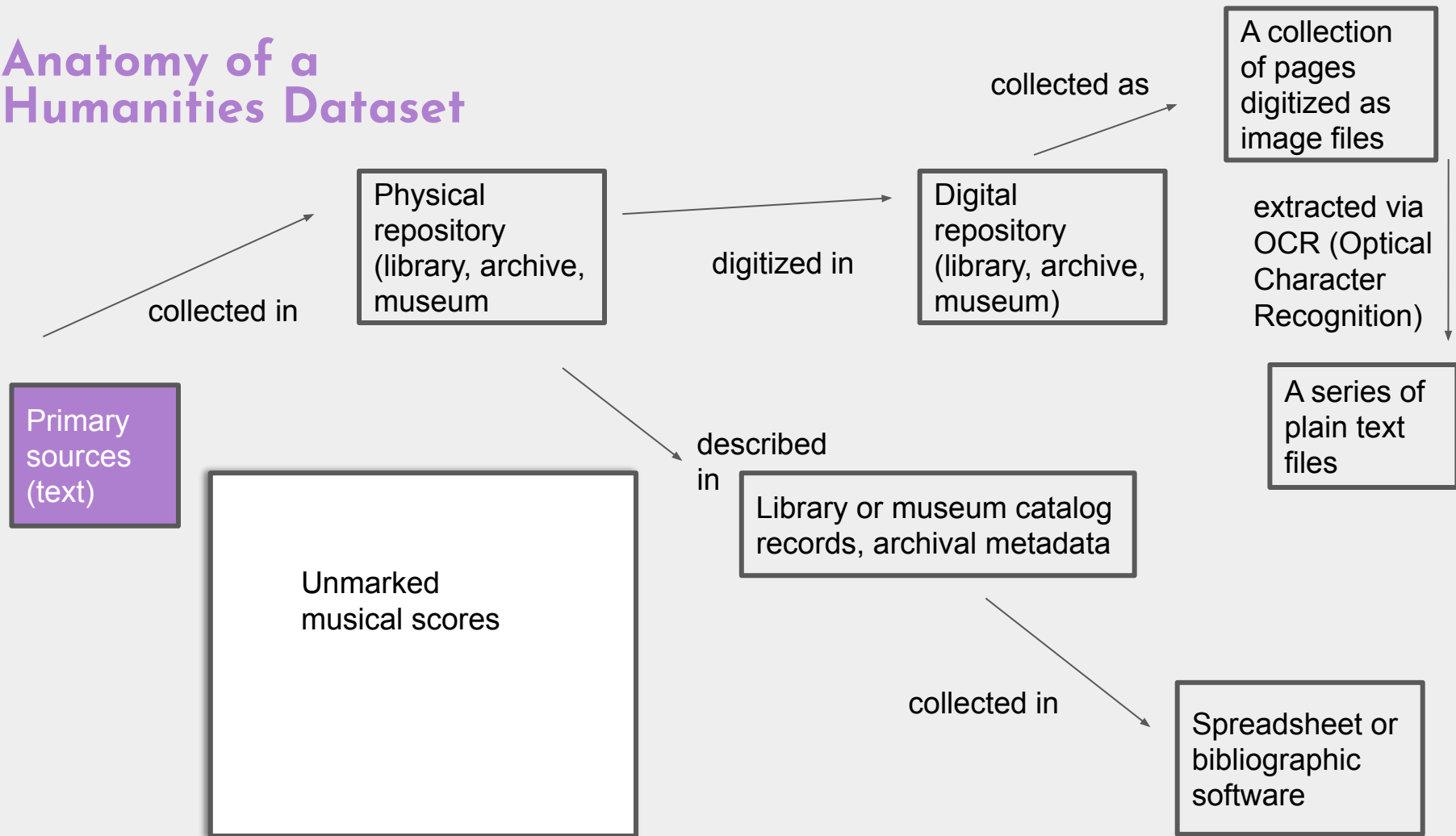
Anatomy of a Humanities Dataset



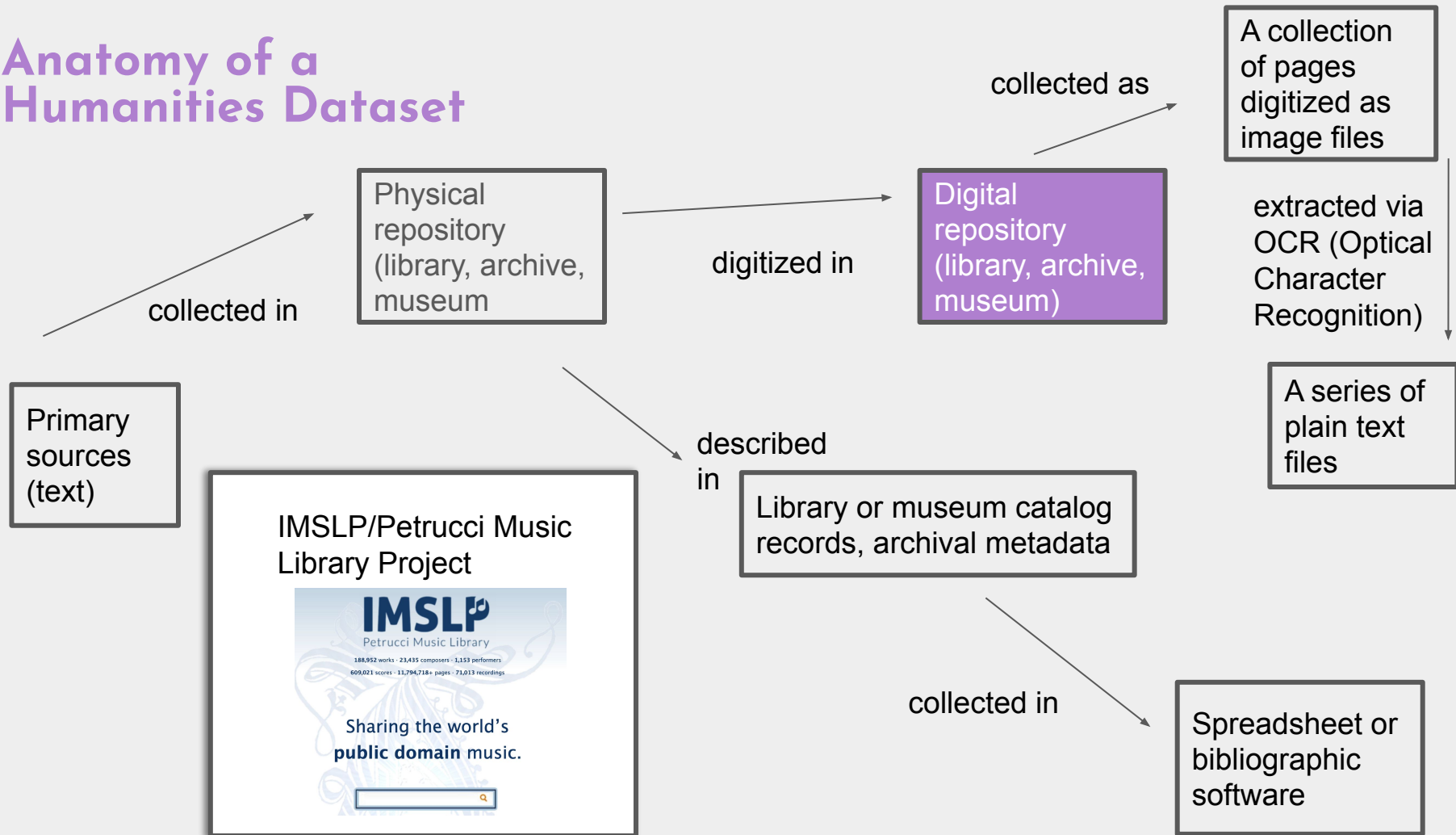
Anatomy of a Humanities Dataset



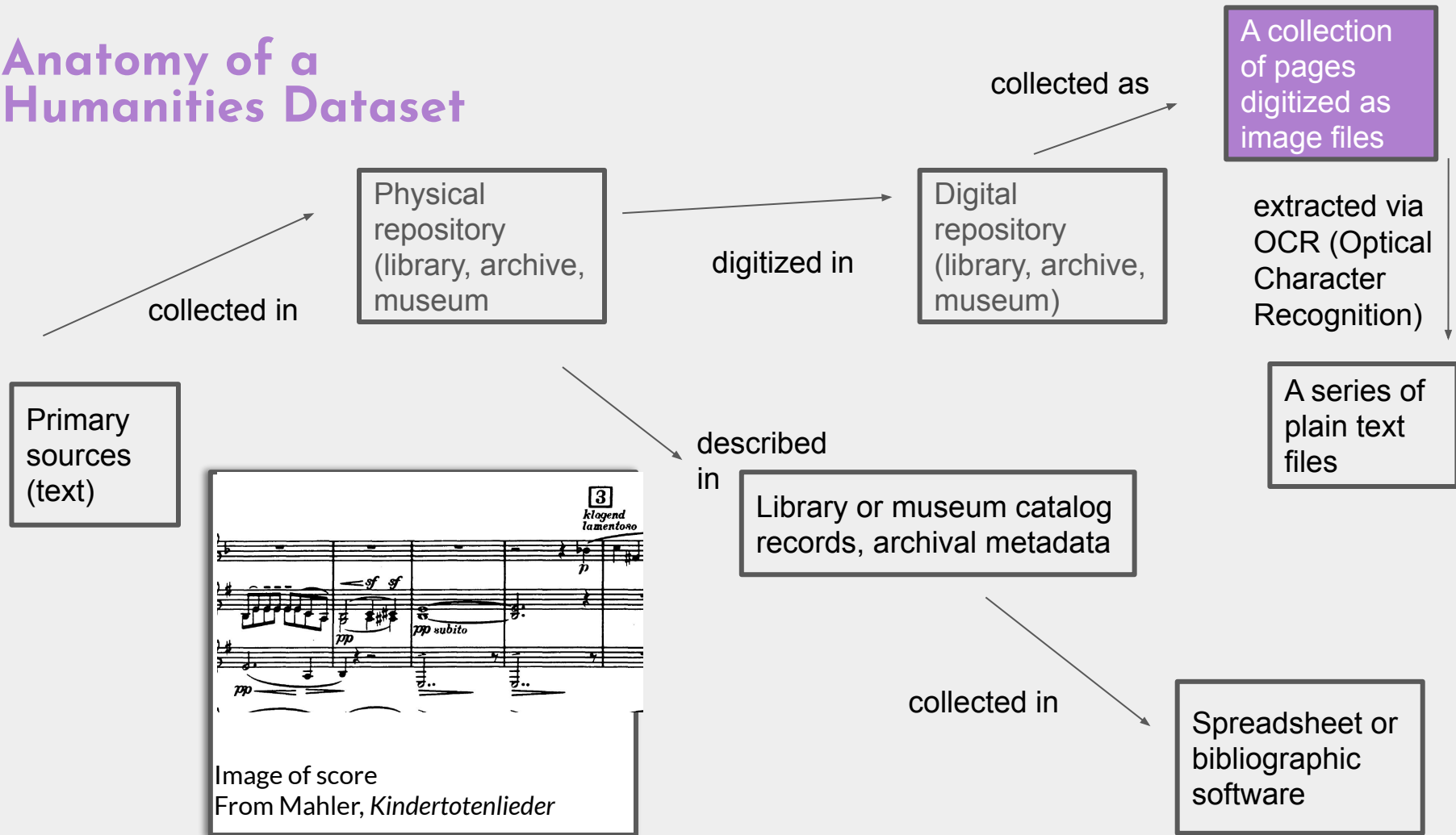
Anatomy of a Humanities Dataset



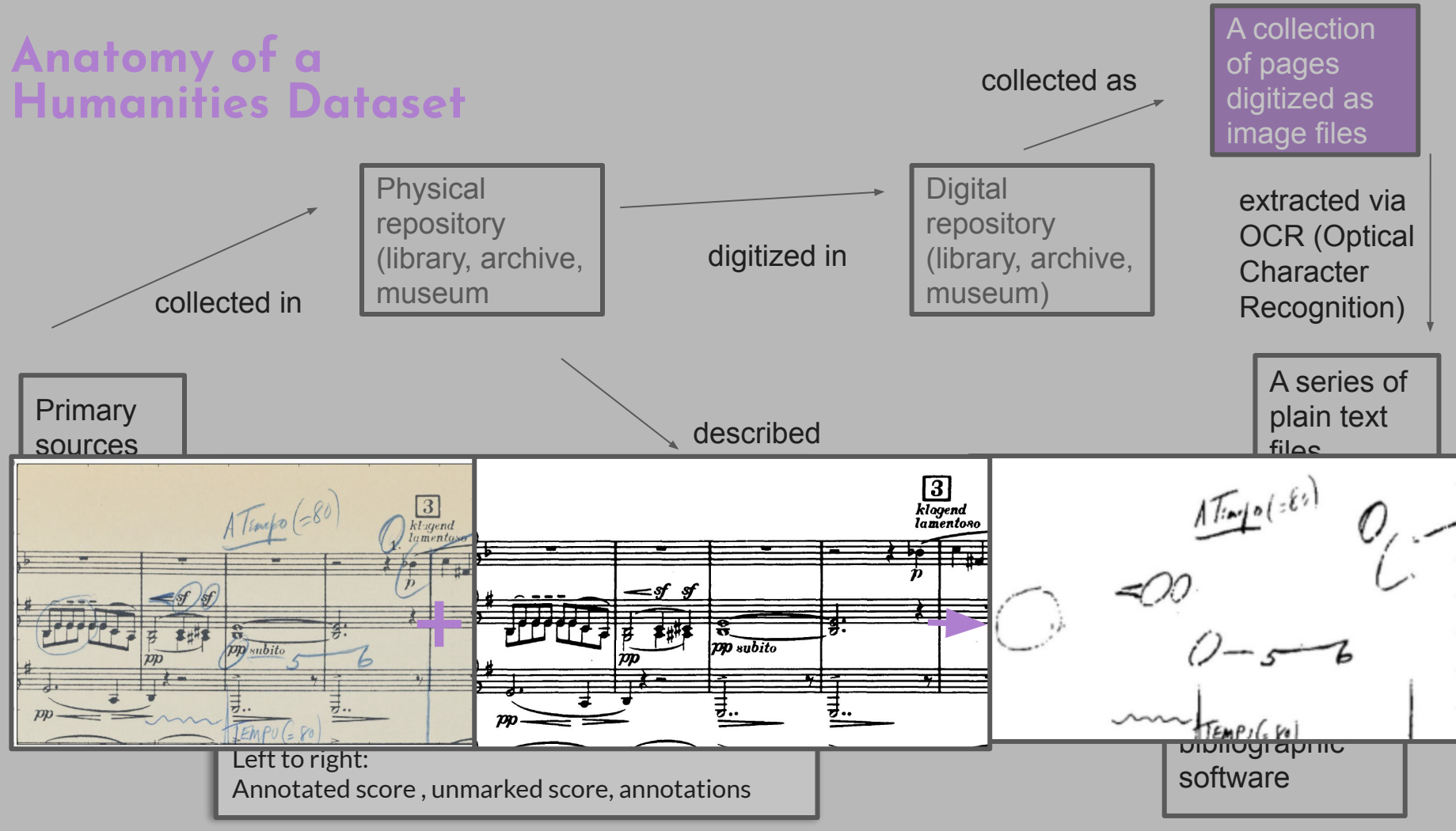
Anatomy of a Humanities Dataset



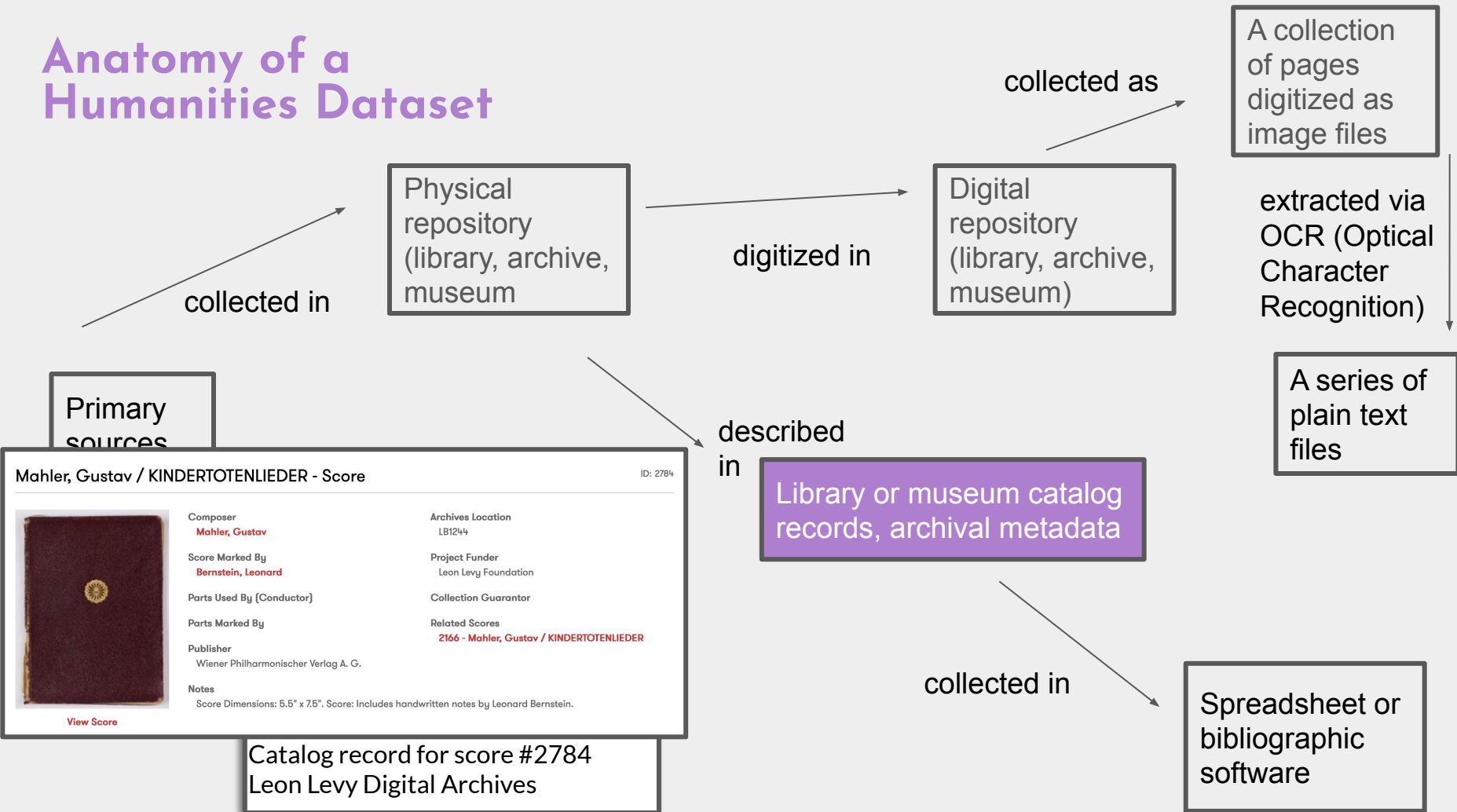
Anatomy of a Humanities Dataset



Anatomy of a Humanities Dataset



Anatomy of a Humanities Dataset



ARCHIVES

What do we mean by mean by
“archive”?

ARCHIVES

Archive as in institutional record repository?

Archive as in “The Papers of So and So”?

Archive as in “Right Click -> add to Archive”

Archive as in web archive?

Archive as in digital archive?

Archive as in *The Archive*?

Trevor Owen, “What Do you Mean by Archive? Genres of Usage for Digital Preservers,” *The Signal - Library of Congress*, (blog), Library of Congress, February 27, 2014

Think back to your sources
from Exercise #2.

- ❖ What's your **archive**?
- ❖ What's your data?
- ❖ What metadata is of
interest to you?

4.

What does a humanities
approach to data look
like?

All datasets are situated

Your data shapes your
research questions

How you define your data
and your sources shapes
your research

5.

Q&A: Ask Us Anything

THANK YOU!

Join us on **Nov. 3** for

Workshop 2: Finding Humanities Data

sceckert@princeton.edu

Readings & References

- Drucker, Johanna. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 005, no. 1 (March 10, 2011). <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>
- Posner, Miriam. "Humanities Data: A Necessary Contradiction." *Miriam Posner's Blog* (blog), June 25, 2015. <https://miriamposner.com/blog/humanities-data-a-necessary-contradiction/>.
- Rosenberg, Daniel. "Data before the Fact." In *"Raw Data" Is an Oxymoron*, edited by Lisa Gitelman and Virginia Jackson, 15–40. Cambridge, MA: MIT Press, 2013. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6462162>
- Underwood, Ted. "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago." *Representations* 127, no. 1 (August 1, 2014): 64–72. <https://doi.org/10.1525/rep.2014.127.1.64>.
- "What Is Humanities Research Data? • Digital Scholarship Services • Lafayette College." <https://dss.lafayette.edu/what-is-humanities-research-data/>
- Wythoff, Grant. "CDH Dataset Curation Orientation," Center for Digital Humanities, June 5, 2019.