
2026-02-25

Data in the Humanities

Wk. 05: Is “Digital Humanities” Now Just... “Humanities”?

Today’s Plan

1. 📄 Discuss final project format
2. 📖 Discuss readings on DH & state of quant methods in humanities
3. 🖋️ Intro to finding and assessing humanities data

<https://gwijthoff.github.io/CDH507/final/>

Readings

Read a chapter relevant to your field from one volume in the **Debates in DH series** (University of Minnesota Press).

Matthew Wilkens, “What Instagram and Community Colleges Tell Us about the Future of Digital Humanities,” *American Literary History* 37, no. 4 (2025).

Leif Weatherby, *Language Machines: Cultural AI and the End of Remainder Humanism* (Univ. Minnesota Press, 2025), Introduction & Conclusion.

Underwood, McGrath, So, & Wellmon, “Culture, Theory, Data: An Introduction” to special issue, *New Literary History* 53, no. 4 (Winter 2023).

Finding & Understanding Data

d, [wikidata_work_qid](#), title, creator, released, year_set, years_elapsed, multiyears, medium, genres, silence_2020, [Q100333537](#), The Silence, *Don DeLillo*, 2020, 2022, 2, , **fiction**, apocalyptic novel, [http](#) [dzilla_2021](#), [Q100977481](#), Godzilla Singular Point, *Toh Enjoe*, 2021, 2030, 9, , **television**, [https://e](#) [_venus_1989](#), [Q101134724](#), Venus Wars, *Yoshikazu Yasuhiko*, 1989, 2089, 100, , **film**, [https://en.wikipe](#) [_2019](#), [Q101208267](#), Dry Drowning, *Giacomo Masi*, 2019, 2030, 11, , **fiction**, visual novel, [https://en.wi](#) [wboy_1999](#), [Q101244908](#), Cowboy Bebop, *TV Tokyo*, 1999, 2021, 22, , **television**, Anime series, [https://en](#) [expanse_2003](#), [Q1027348](#), The Expanse, *Allan Kroeker*, 2003, 2153, 150, , **fiction**, science fiction, [http](#) [_exterminators_1983](#), [Q10290084](#), Exterminators of the Year 3000, *Giuliano Carnimeo*, 1983, 3000, 10 [anet_1973](#), [Q1029967](#), Planet of the Daleks, *Terrance Dicks*, 1973, 2500, 527, , **fiction**, [https://en.w](#) [uture_1902](#), [Q10303294](#), Future of a New China, *Liang Qichao*, 1902, 1962, 60, 1903–1962, **fiction**, Nove [_hunger_2015](#), [Q10307713](#), The Hunger Games: Mockingjay ▣ Part 2, *Francis Lawrence*, 2015, 2093, 78, [fast_1991](#), [Q1031248](#), ""Fast Forward"" (from The Flash)", *Warner Bros. Television Studios*, 1991 [e_hellraiser_1996](#), [Q1031262](#), Hellraiser: Bloodline, *Kevin Yagher; Alan Smithee; Joe Chappelle*, 1982, [Q103411](#), V für Vendetta, *Alan Moore*, 1982, 1997, 15, , **comics**, Comic, [https://en.wikipedia.org/](#) [berpunk_1990](#), [Q1034070](#), Cyberpunk 2020, *R. Talsorian Games*, 1990, 2020, 30, , **misc**, Board game, [https](#) [2001_1968](#), [Q103474](#), 2001: A Space Odyssey, *Stanley Kubrick*, 1968, 2001, 33, , **film**, [https://en.wiki](#) [_captain_1967](#), [Q1035150](#), Captain Scarlet and the Mysterons, *Gerry Anderson*, 1967, 2068, 101, , **tele** [ien_1979](#), [Q103569](#), Alien, *Ridley Scott*, 1979, 2122, 143, , **film**, [https://en.wikipedia.org/wiki/Alie](#) [uild_2007](#), [Q10357938](#), Rebuild of Evangelion, *Hideaki Anno*, 2007, 2015, 8, , **film**, Film series, [https:](#) [transcendence_2014](#), [Q10384115](#), Transcendence, *Wally Pfister*, 2014, 2021, 7, , **film**, [https://en.wiki](#) [2007](#), [Q10394592](#), Year Zero, *42 Entertainment*, 2007, 2022, 15, , **video game**, [https://en.wikipedia.or](#) [_tribes_2021](#), [Q104213753](#), Tribes of Europa, *Florian Baxmeyer*, 2021, 2074, 53, , **television**, science [_outside_2021](#), [Q104241071](#), Outside the Wire, *Mikael Håfström*, 2021, 2036, 15, , **film**, war film, [https](#) [dowrun_1989](#), [Q1042442](#), Shadowrun, *Catalyst Game Labs*, 1989, 2021, 32, , **misc**, Roleplaying game, [https](#) [pace_1975](#), [Q1045743](#), Space: 1999, *Ray Austin*, 1975, 1999, 24, 1999–2006, **television**, TV series, [https](#) [aliens_1986](#), [Q104814](#), Aliens, *James Cameron*, 1986, 2179, 193, , **film**, [https://en.wikipedia.org/wiki](#) [asshern_2004](#), [Q1048319](#), Casshern, *Kazuaki Kiriya*, 2004, 2075, 71, , **film**, [https://en.wikipedia.org/](#) [lame_2019](#), [Q104856357](#), Blame the Hero, *Brandon Rogers*, 2019, 2033, 14, , **fiction**, [https://en.wikipe](#)

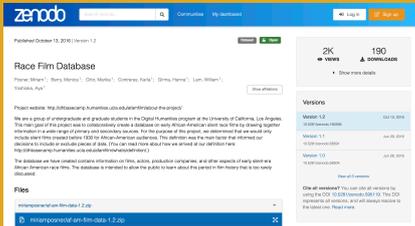
```
wikidata_ID, title, creator, released
Q29649659, Too Like the Lightning, Ada Palmer, 2016
```

	wikidata_ID	title	creator	released
1				
2	Q29649659	Too Like the Lightning	Ada Palmer	2016

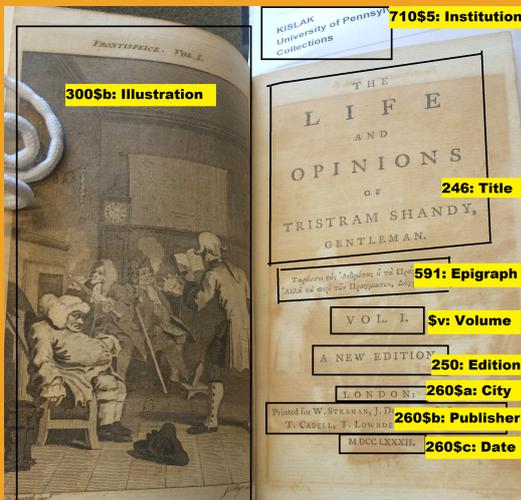
unstructured data
semi-structured data
structured data (CSV / JSON)

Fields in Posner et al.'s dataset:

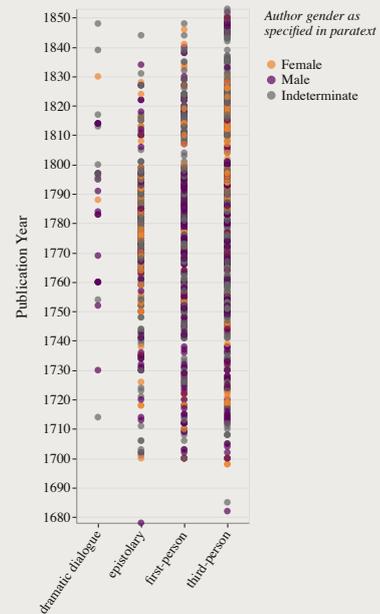
- Title
- Genre
- Primary Source Verification
- Production Company
- Production date
- Distribution date
- Length (reels)
- Actors
- Directors
- Producers
- Camerawork
- Source
- Writer
- Otherwise Involved
- Link to media
- Distributor



Race Film Database (UCLA) & data paper



Narrative Voices in Early Novels



Early Novels Dataset

“Tidy Data”

Hadley Wickham, 2014

- one variable per column,
- one observation per row
- consistent data types
- unique identifiers for each object
- clear names for your fields
- *if you find you’re trying to fit too much information into a column or row, use another table and cross-reference*

FAIR Data

articulated 2016

Findable
Accessible
Interoperable
Reusable

Linked open data

Data that is “structured using recognized standards so that computers interrogating the data can process it consistently,” that refers to people / places / things using the same identifiers, and that is published openly.

Controlled vocabularies are used to:

→ Internally link or identify records within datasets

→ Link to institutionalized standards that other scholars / professionals might use (“Linked Open Data”)

Examples

[The Getty Vocabularies](#)

[Oxford Dictionary of National Biography](#)

[VIAF: The Virtual International Authority File](#)

[MIMO Thesaurus of Musical Instrument Names](#)

The Programming Historian's [intro to linked data](#)

sources for humanities data

[Responsible Datasets in Context](#), eds. Fernandez, Posner, Preus, Singh, & Walsh

[Post45 Data Collective](#), eds. Walsh, Manshel, Porter

Amanda Wyatt Visconti. “[Interesting digital humanities data sources](#),” UVA Scholars Lab Blog, (August 26, 2025).

Matt Lavin’s [Humanities Data](#)

[Journal of Open Humanities Data](#)

Jeremy Singer-Vine’s [Data Is Plural](#) archive

The Pudding's [data repository](#)

[CDH-curated list of humanities datasets](#)

do this:

- Find a CSV, download it, open it in a spreadsheet program (Excel, Numbers, Google Sheets).

answer this:

- What does your dataset contain?
- How does it uniquely identify each record?
- How does it format people / places / dates?
- What categories and values are over-represented?
- Do you see any initial patterns or trends?