

FAIR enough:
Understanding the repository as a
route forward in the publication of
Humanities and Cultural Heritage data

Daniel Paul O'Donnell

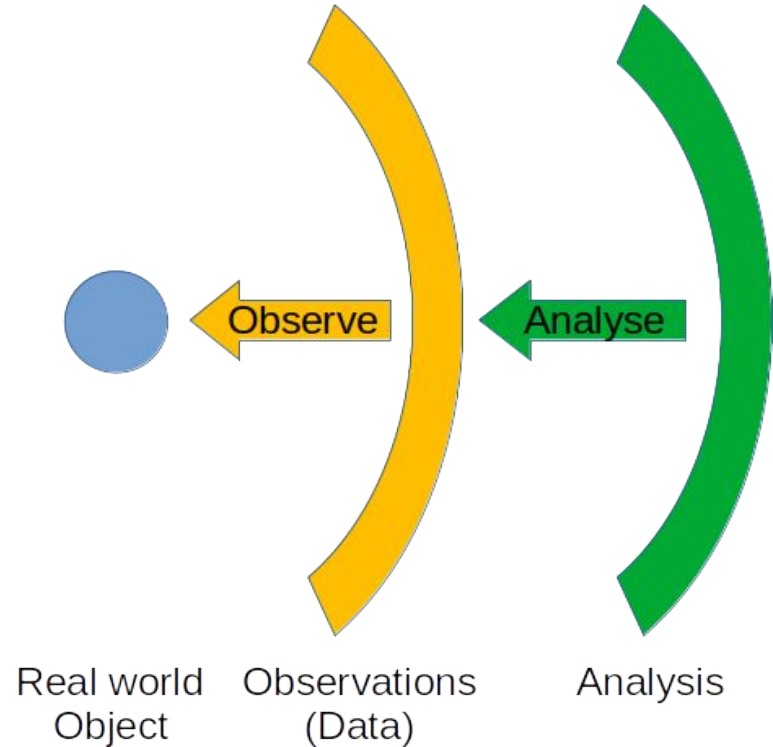
FSCI 2020 Online Lightning Talks
August 3, 2020

The argument

- Current repository structures are not optimal for the publication, citation, and reuse of Humanities and Cultural Heritage (H/CH) Data
- This is because
 - H/CH data is fundamentally different in form and use than scientific data
 - Repositories have not been purposeful in attempting to accommodate the H/CH use case
- With very little work, they can be adapted
- FAIR data systems would mark a great improvement over traditional models.

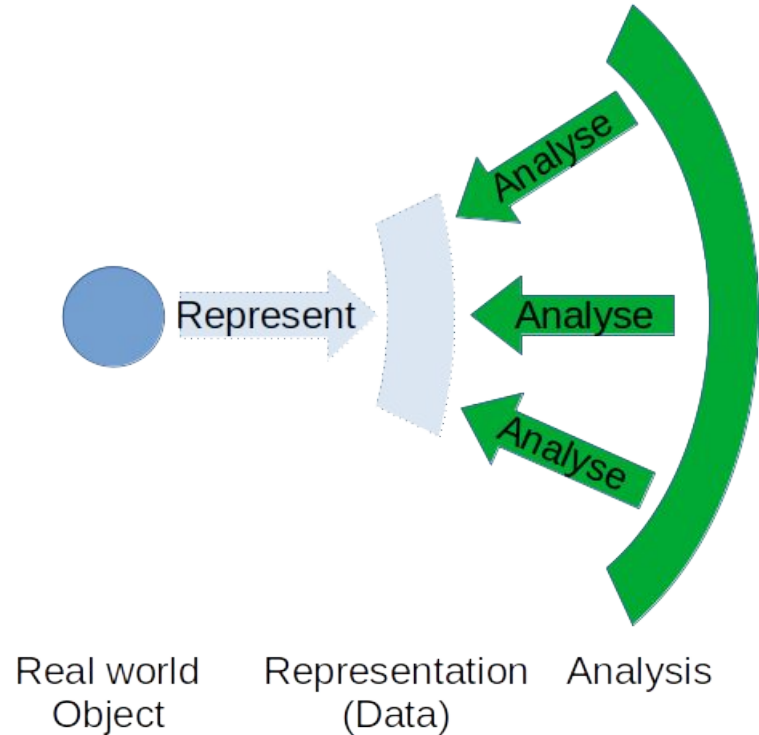
Humanities/Cultural Heritage vs. Scientific Data

- In other domains, “data” (“given things”) is more properly “capta” (“taken”): generated through experiment, observation, and measurement, then observed
- Think about Darwin and his work in the Galapagos Islands
 - What is his data?
Finches? Notes?
- Repositories allow the sharing of the intermediate layer.



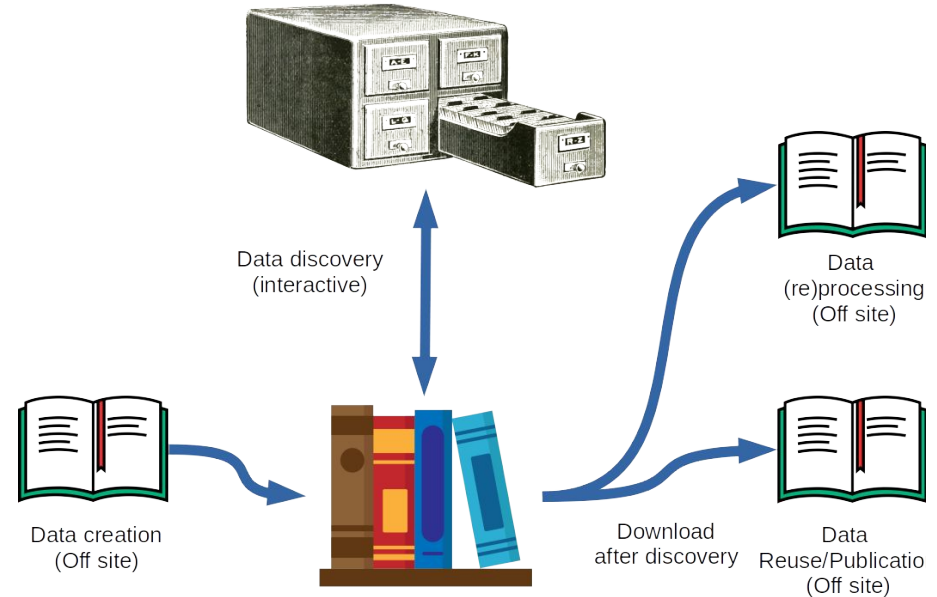
Humanities/Cultural Heritage vs. Scientific Data

- H/CH is “data” not “capta”
- “Small, Thick, and Slow”
 - **Small:** Representational rather than experimental (Finches, **Notes**);
 - **Thick:** Very detailed curation (100k on 9 lines);
 - **Slow:** Frequently reanalysed (Jane Austen Studies)
- Fundamental purpose is reanalysis



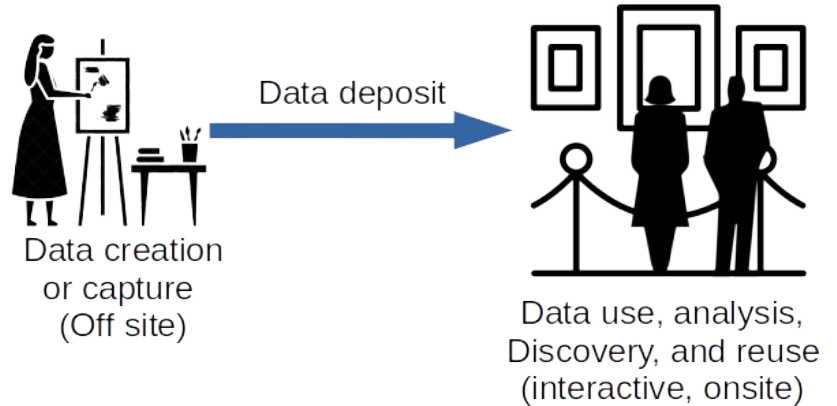
Current Repository Model is a Library

- Data is “composed” offsite
- Catalogued and “shelved” in Repository
- “Checked out” (Downloaded to separate system for analysis, reuse, etc.)



Humanities ideal would be more like a gallery

- Data “composed” off-site
- Catalogued and “displayed” onsite
- Discovered, used, analysed, and reused on-site
- Everything we need for this already exists:
 - Streaming (Figshare)
 - Restful/PID-based URLs
 - Typed LOD capabilities (Zenodo)



Let's think about H/CH
reuse as purposefully
as we have Scientific!