



Corpus Enrichment Group

Preamble

We've got a lot of data - pretty much everything published in English in the eighteenth century - but it's messy. We need your help to clean it up, develop powerful new search filters, reconcile metadata with Wiki data entities, and key the search results to our library holdings. Coders are welcome, as is anyone with a hankering to dig into some library metadata!

Main Goal

There are two main goals for this group. The first is to enrich our metadata by reconciling named entities in the ECCO metadata with online data repositories, possibly including HOLLIS metadata, Wikidata, Geonames, and WorldCat. The second is to clean up the bad OCR for ECCO by running it against a standard dictionary. Better data, not perfect data, is the goal!

Other goals / instructions (in order of priority)

1. Combine named author entities with *DNB* data (biography, likenesses, archival holdings).
2. Clean and mount other corpora (EEBO, HathiTrust)
3. Work with Research Exploration group to develop queries and search filters
4. Identify redundant author names and reduce to single entities

GitHub Contents (may be added to before and during the event):

- A snippet of the full metadata for Eighteenth Century Collections Online
- Link to Underwood Git for cleaning ECCO metadata
- List of possible search filters
- Group Notebook
- Code Uploads

Other groups' tasks that might be of interest

- **Research Exploration** (for useful metadata to acquire, in order to make robust search filters);
- **Search Development**
- **Data Visualization and Analytic Tools:** Your metadata will provide the basis for their tools