EASY

* Norton anthology of literature index info

INTERMEDIATE

* College reading lists
  + “survey of english literature”
* Public school reading lists
  + Should contain, titles, authors
* Wikipedia articles & relative presence?
  + List of literary works that are most referenced / longest / have the most citations?
* Aggregated datasets from other similar research projects?
  + What other researchers have found can be used to characterize the lit canon
* Amazon reviews?
* Library route?
  + Banned book lists?
* English literature anthologies
* Google related search results?
  + “survey of English literature”
  + “literary canon”
* Digitized text repository catalogue data?

HARD

* Theory to guide dataset selection?
  + Lots of differing opinions on what the canon is (are there several canons?)

NEED:

* I need more information about theory to guide my dataset collection

REPOSITORIES:

**UK RED**: UK RED is an open-access database housed at The Open University containing over 30,000 easily searchable records documenting the history of reading in Britain from 1450 to 1945. <http://www.open.ac.uk/Arts/reading/UK/explore.php>

**Oxford Text Archive**: The OTA collects, catalogues, preserves and distributes high-quality digital resources for research and teaching.

**OCLC catalogue** <https://www.oclc.org/developer/develop/web-services/worldcat-search-api.en.html>

API:

Wikidata API:

plan of action:

- explore datasets you can use to map author dates against author names (to weed out authors who couldn't be in johnson's dictionary or OED) (wikidata.org? encyclopedia britannica?)