Data Science with Open Book Metadata

Micah Altman
MIT Libraries
Center for Equitable and Open Scholarship (CREOS)
good.

not good enough.

This talks reflects on a six-month project exploring the evidence base on open monographs.

More data from and about scholarship is openly available than ever.

And coverage of subjects, people, behaviors, languages, and forms of knowledge remains narrow.
How has open monograph publication changed?
much growth. lower fees.
Directory of Open Access Books (DOAB), provides metadata on tens of thousands of individual monographs.

OpenAPC provides title-level data on processing charges, supplied by a number of consortial initiatives.

OpenBookPublishers provides title-level usage data on the titles it publishes.

OpenLibrary can be used to obtain additional descriptive metadata for titles, including subject headings.


I40C initiative. Provides citation data for books with DOI.s

HathiTrust and JSTOR provide public metadata, rights metadata, and text-analytic metadata (e.g. ngrams) for the (open) books in their collection.
not enough...

unique global persistent identifiers

metadata quality

robust API clients

collection coverage
What’s missing …

Durable & reliable content access

Ecosystem health indicators

Costs, fees, and purchases - economics

Participant demographics

Accessibility metadata

Verifiable rights information


https://xkcd.com/523/
Want to see more?

Visit MIT Libraries Center for Equitable and Open Scholarship (CREOS)

Center for Research on Equitable and Open Scholarship
Advancing knowledge in service of equitable and open scholarship