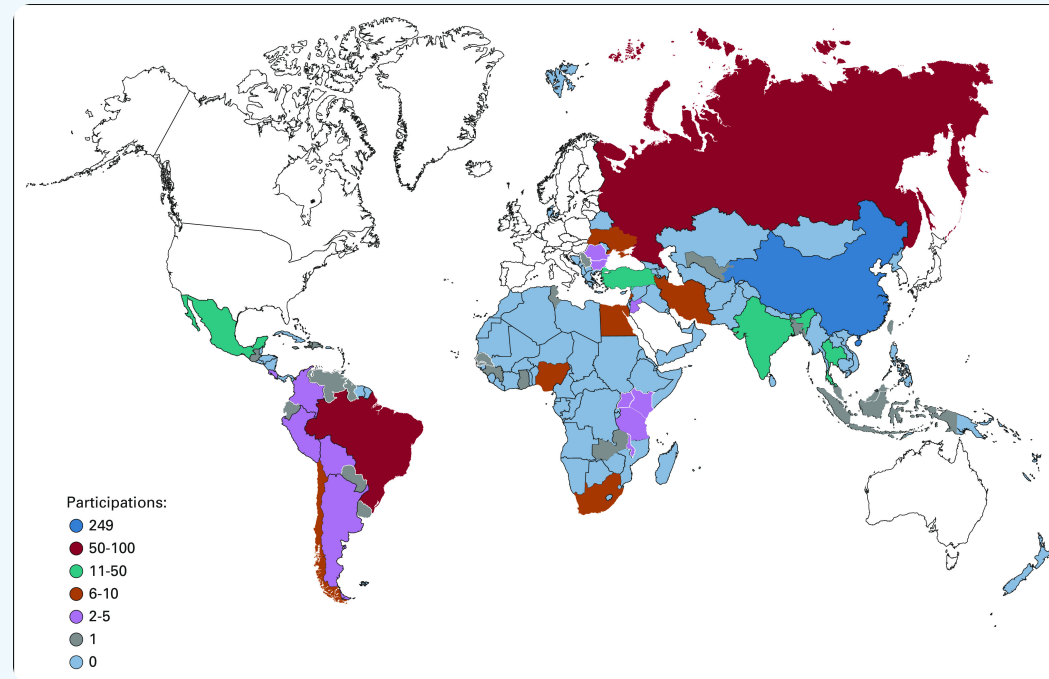


HemOnc.org Q2

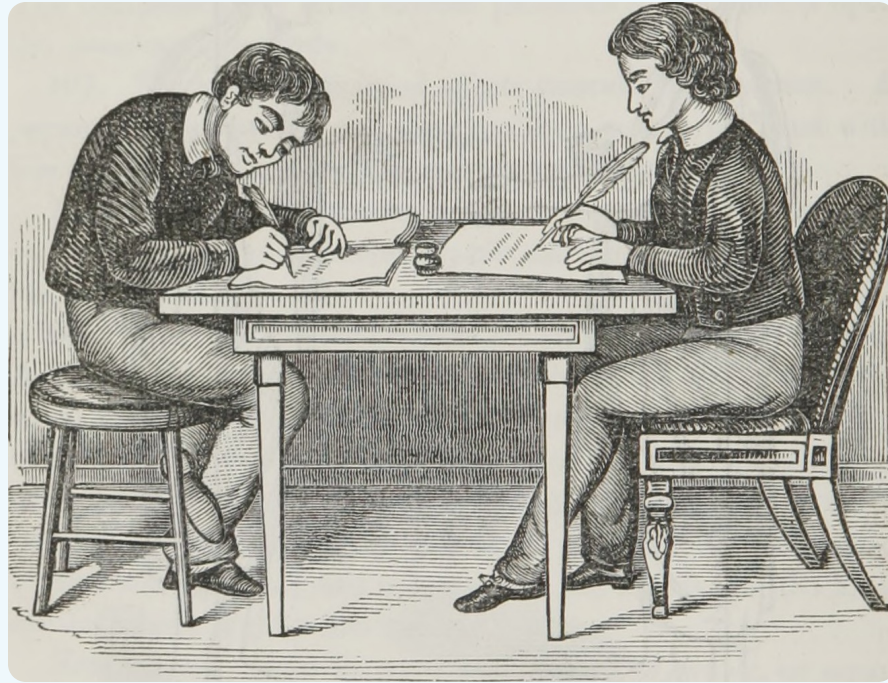
R-CEOP

Where in the world is hematology?



~5000 abstracts reviewed, ~500 had author from LMICs

Where in the world is hematology?



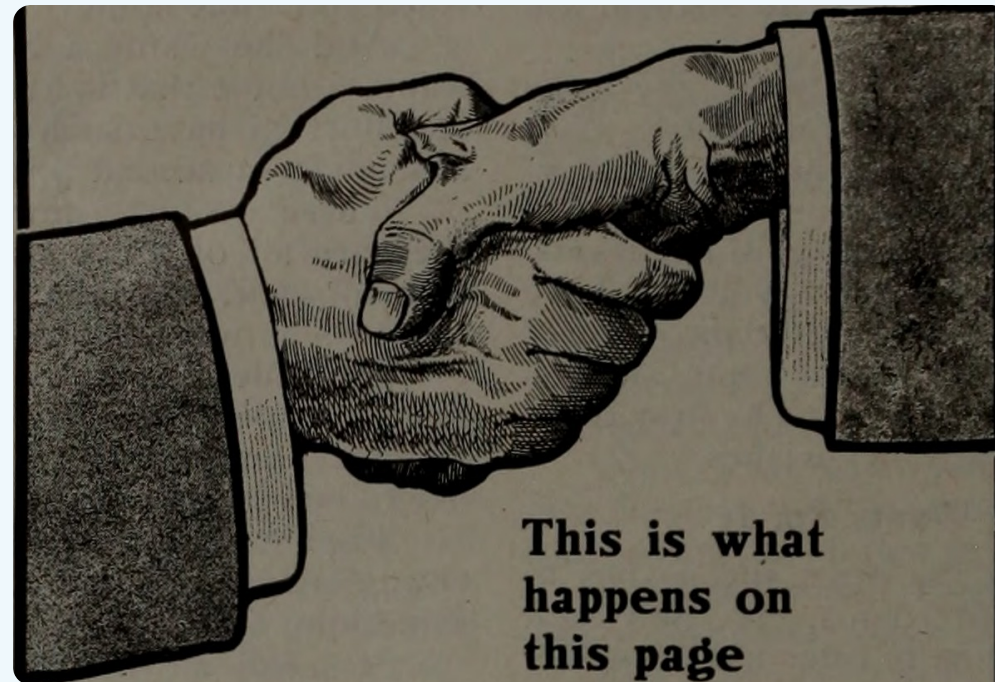
manual extraction is slow and ruins your back

Right [patient, drug, time, dose, route, *place*]

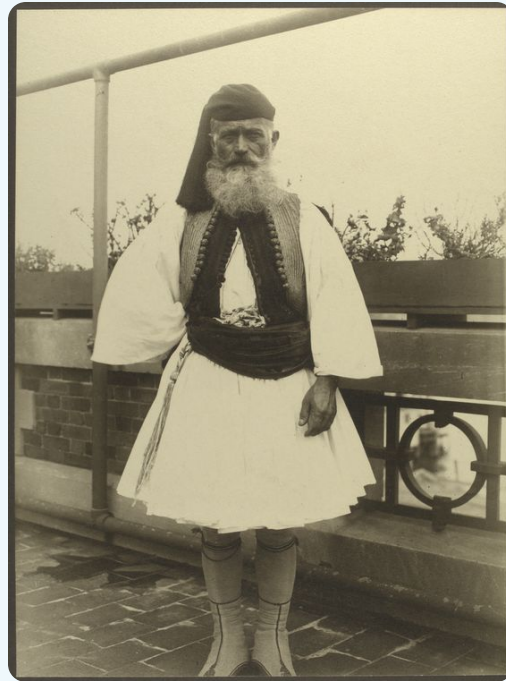


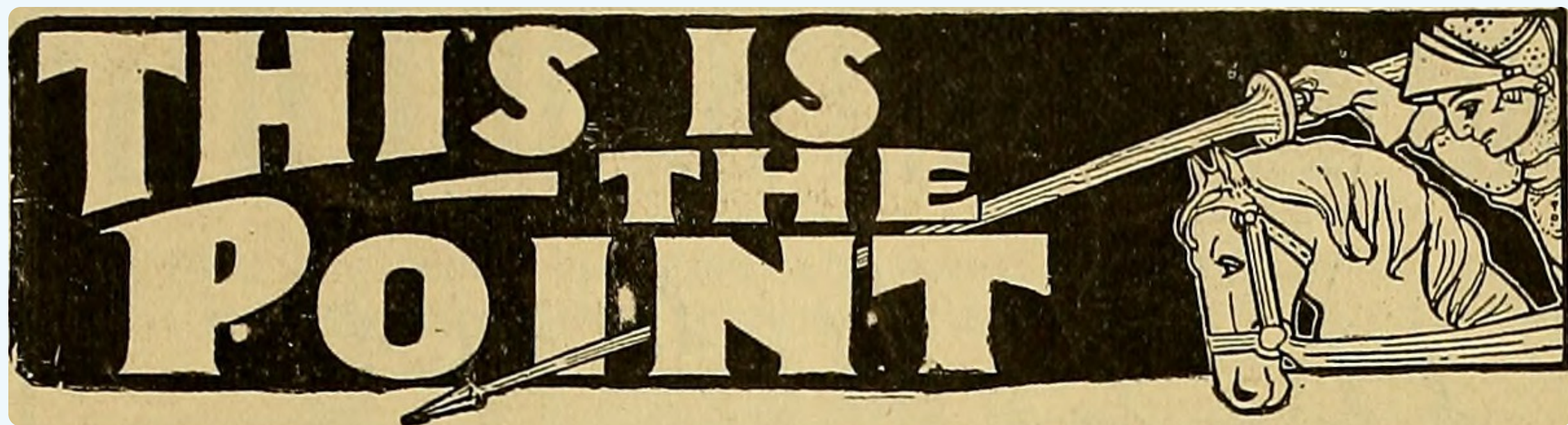
Léon Tshilolo, MD, PhD. ASH 2018, speaking on hydroxyurea in Sub-Saharan Africa.

Myopia

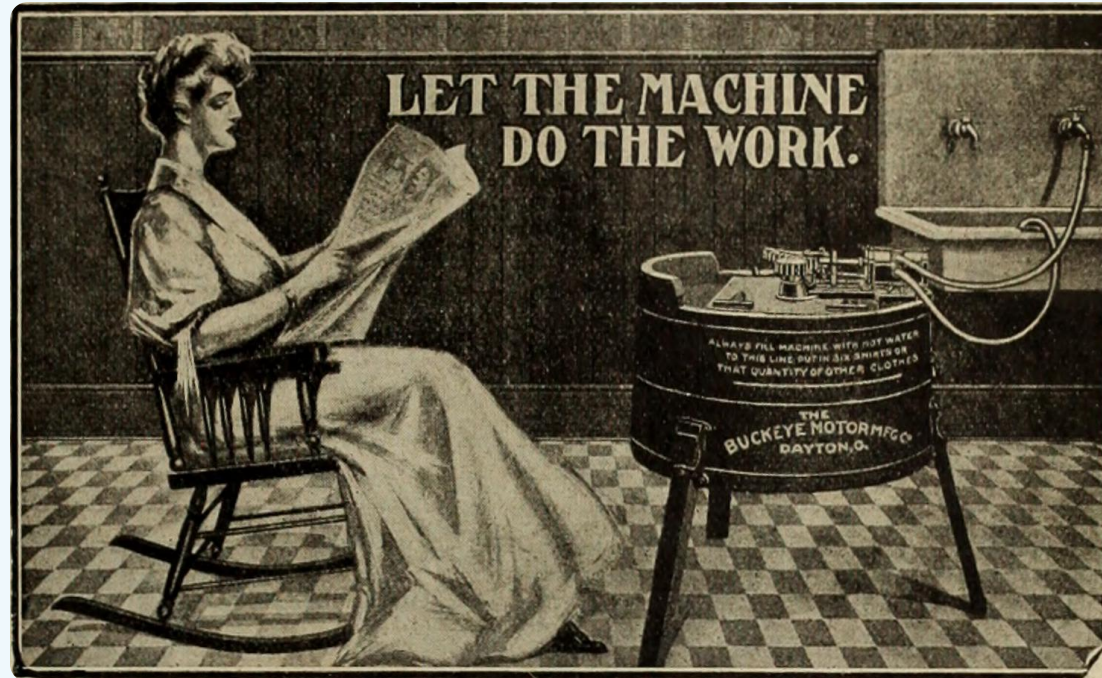


Gatekeeper/Guide



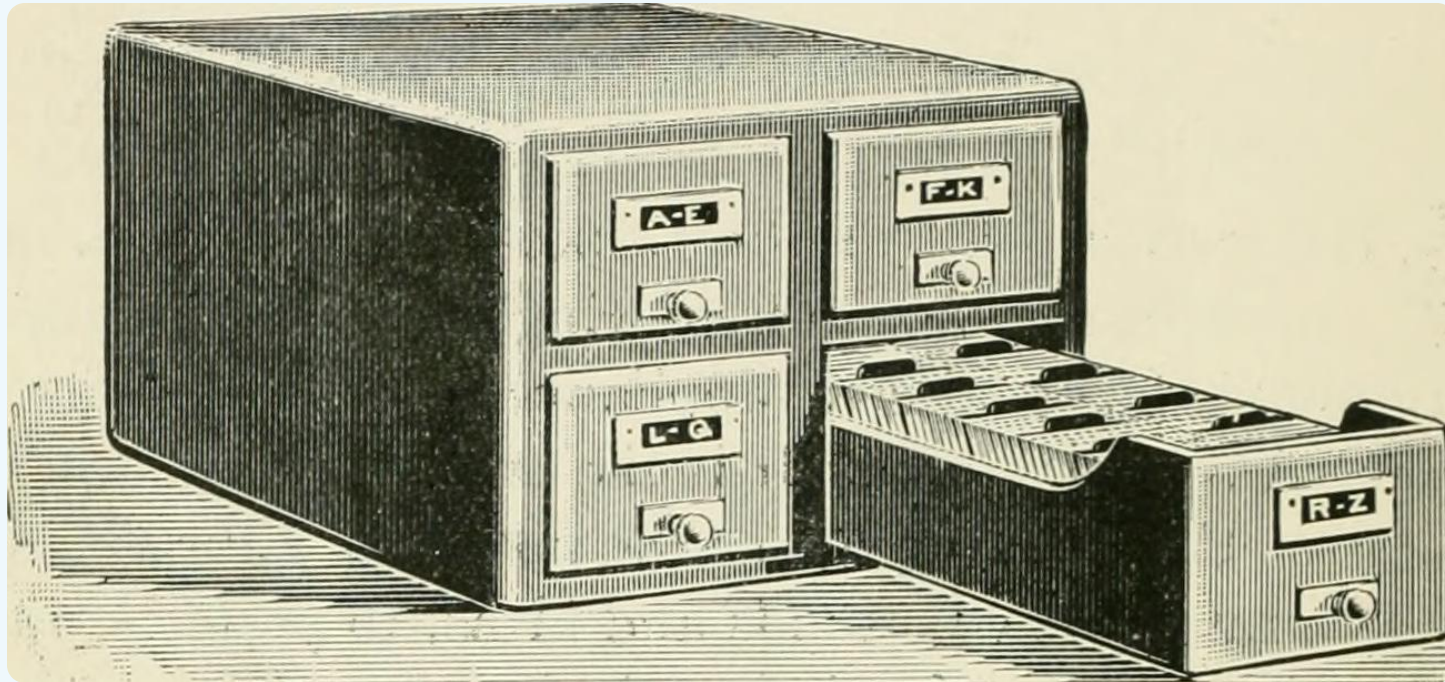


Current project



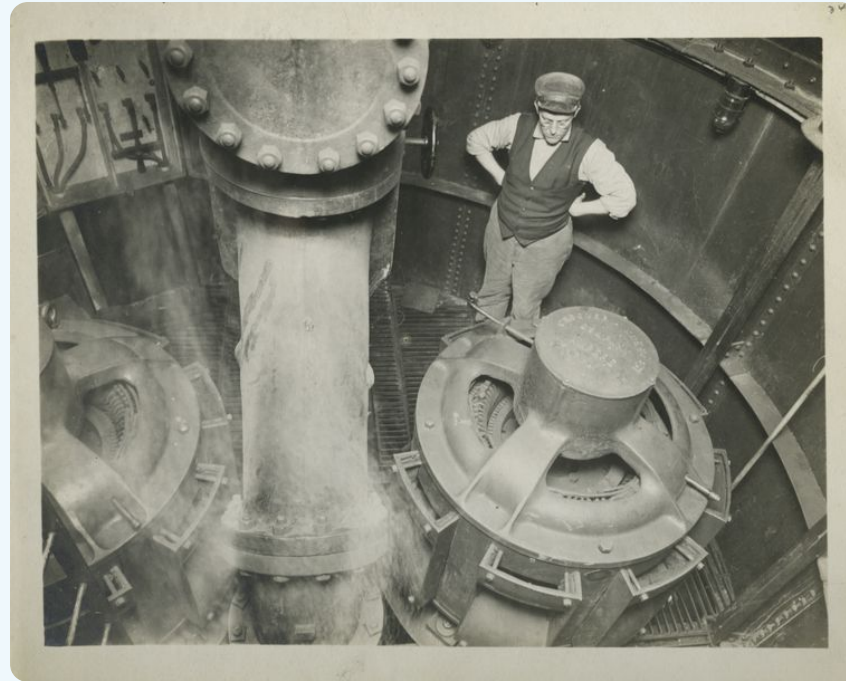
What if we could automate the extraction and structuring of all ~100k abstracts?

Current project



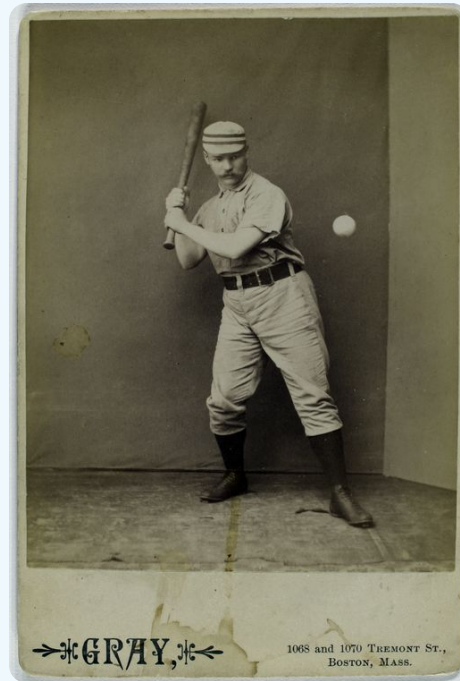
keep it public, keep it safe

Current project



analysis pipeline

Current project



BYOA (bring your own analysis)

Planned analyses

trends

geographic

chronologic

topologic

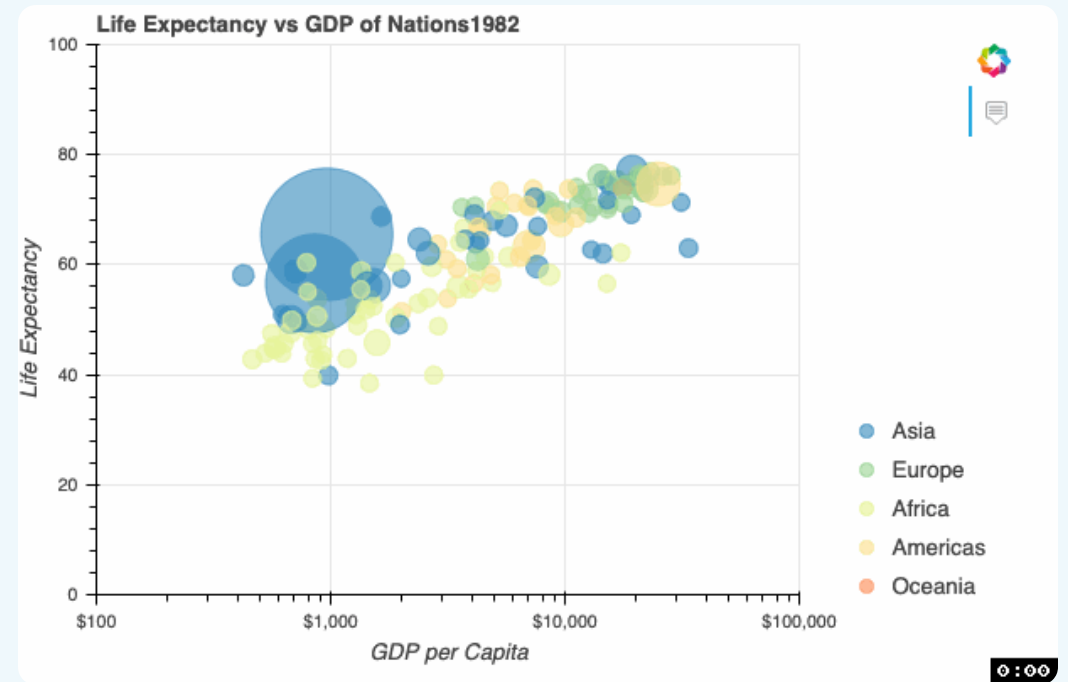
ontologic

biographic

disparities

e.g. overlay maps of the disease
itself (incidence, morbidity, etc.)
and research about that disease

funding/resource allocation



h/t Hans Rosling

Challenges

Structure, standards

<https://schema.datacite.org>

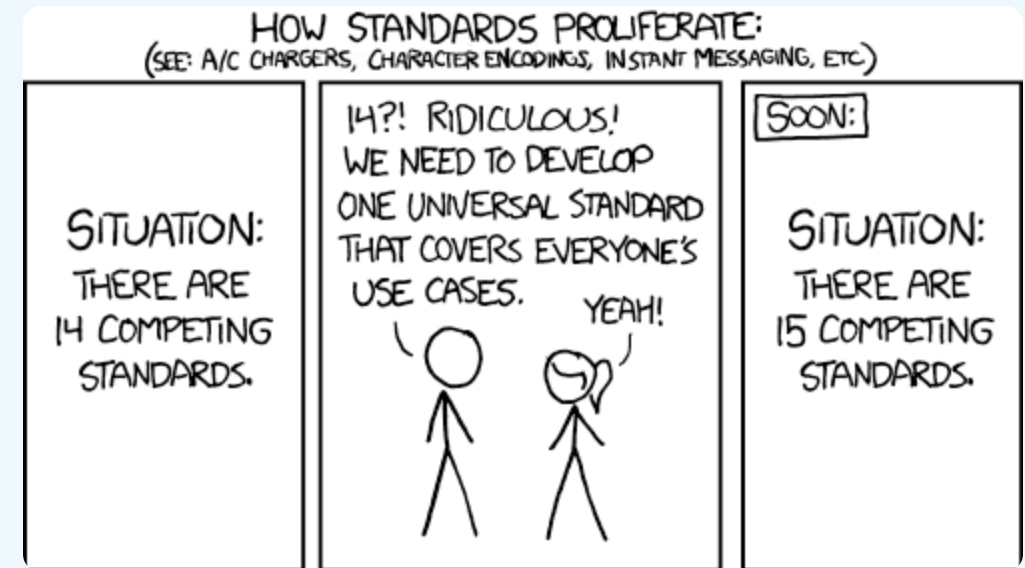
FAIR

Findable

Accessible

Interoperable

Reusable

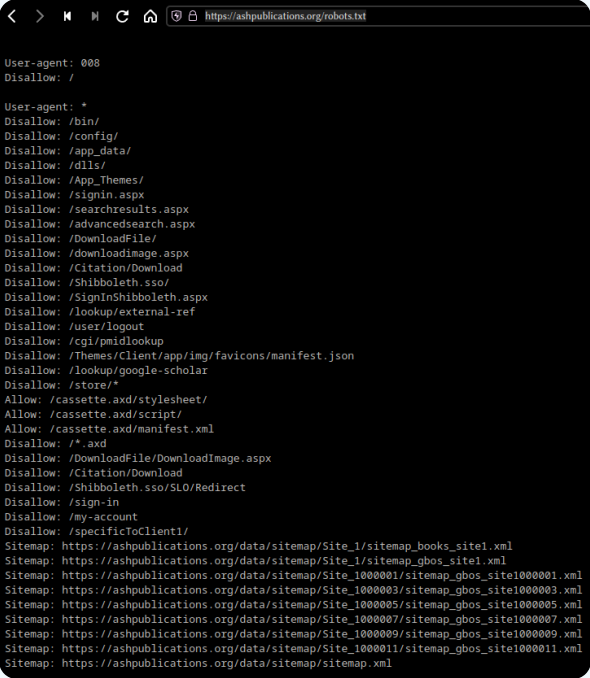


Challenges

robots.txt

<https://ashpublications.org/robots.txt>

cf hiQ vs LinkedIn

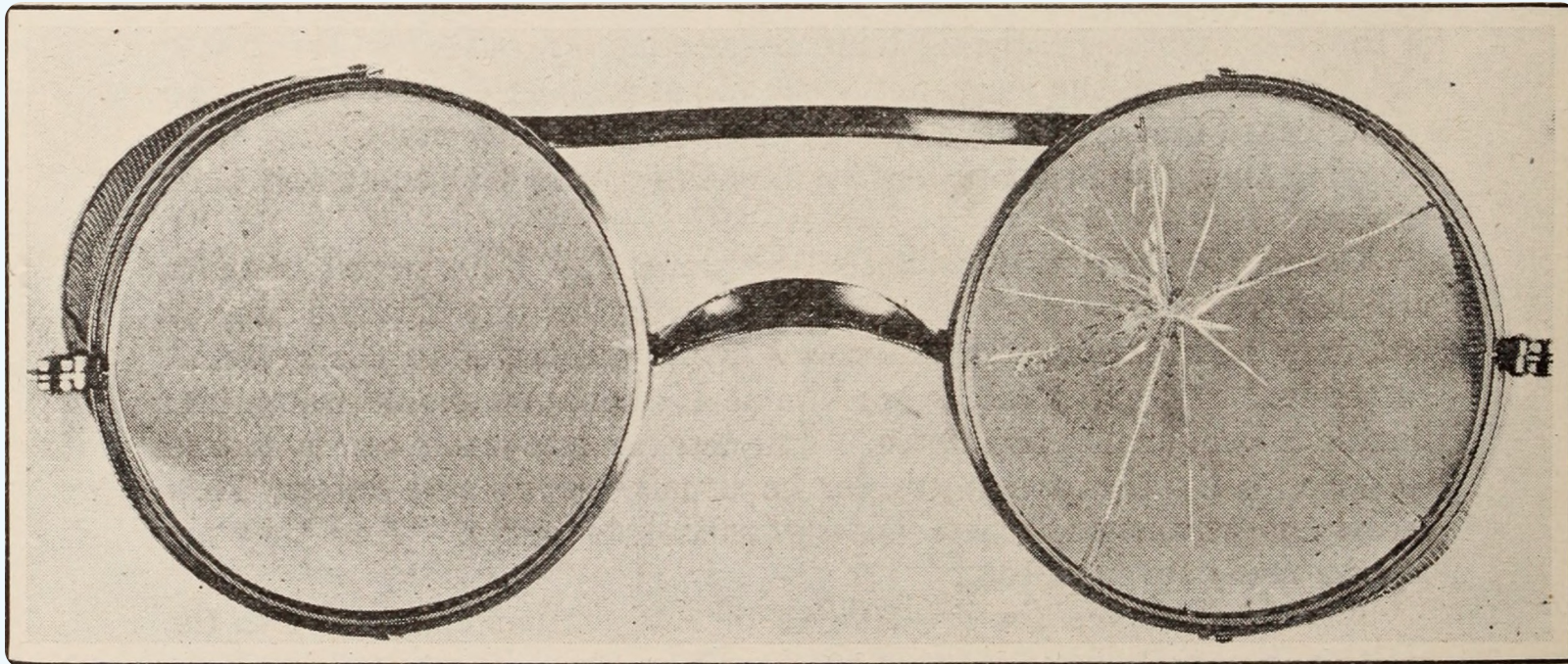


```
< > M C https://ashpublications.org/robots.txt

User-agent: 008
Disallow: /

User-agent: *
Disallow: /bin/
Disallow: /config/
Disallow: /app_data/
Disallow: /dlls/
Disallow: /App_Themes/
Disallow: /signin.aspx
Disallow: /searchresults.aspx
Disallow: /advancedsearch.aspx
Disallow: /DownloadFile/
Disallow: /downloadimage.aspx
Disallow: /Citation/Download
Disallow: /Shibboleth.sso/
Disallow: /SignInShibboleth.aspx
Disallow: /lookup/external-ref
Disallow: /user/logout
Disallow: /cgi/pmidlookup
Disallow: /Themes/Client/app/img/favicons/manifest.json
Disallow: /lookup/google-scholar
Disallow: /store/*
Allow: /cassette.axd/stylesheets/
Allow: /cassette.axd/script/
Allow: /cassette.axd/manifest.xml
Disallow: /*.axd
Disallow: /DownloadFile/DownloadImage.aspx
Disallow: /Citation/Download
Disallow: /Shibboleth.sso/SLO/Redirect
Disallow: /sign-in
Disallow: /my-account
Disallow: /specificToClient1/
Sitemap: https://ashpublications.org/data/sitemap/Site_1/sitemap_books_site1.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1/sitemap_gbos_site1.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000001/sitemap_gbos_site1000001.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000002/sitemap_gbos_site1000002.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000003/sitemap_gbos_site1000003.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000004/sitemap_gbos_site1000004.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000005/sitemap_gbos_site1000005.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000006/sitemap_gbos_site1000006.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000007/sitemap_gbos_site1000007.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000008/sitemap_gbos_site1000008.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000009/sitemap_gbos_site1000009.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000010/sitemap_gbos_site1000010.xml
Sitemap: https://ashpublications.org/data/sitemap/Site_1000011/sitemap_gbos_site1000011.xml
Sitemap: https://ashpublications.org/data/sitemap/sitemap.xml
```

Challenges



missing, incomplete, wrong

Challenges



Web scraping is the bug. Not sure who the other guy is.

Challenges (also a feature)

home / abstracts_small

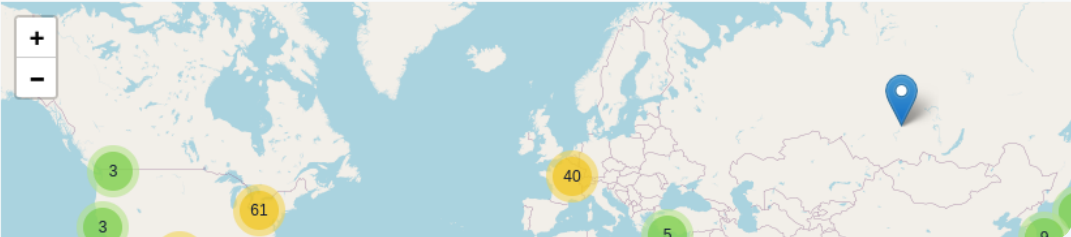
abstracts_small

Custom SQL query returning more than 1,000 rows ([hide](#))

```
1 select doi, article_title, article_date, session_type, topics,
   author_names, author_affiliations, first_author_latitude,
   first_author_longitude from abstracts order by doi limit 5000
```

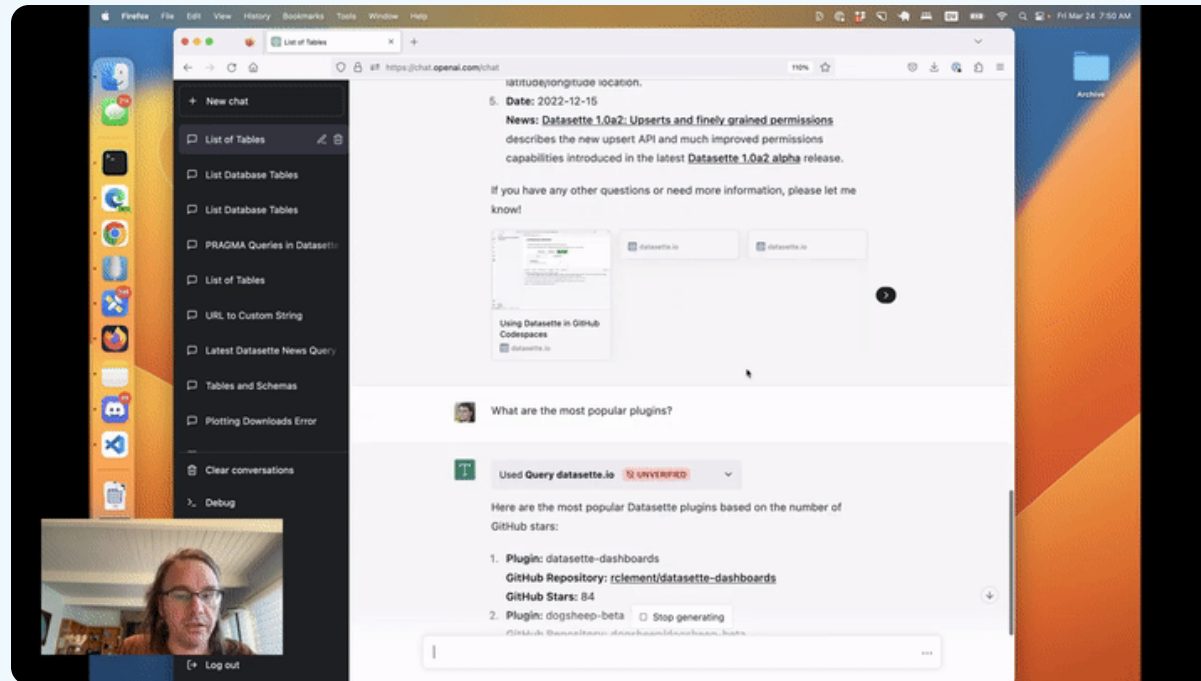
[Format SQL](#) [Run SQL](#)

This data as [json](#), [CSV](#)



If you can write SQL, you can analyze the data.

Challenges (also a feature)



ChatGPT to the rescue?

Show and tell (click here)

