Searching the Deep Web & Grey Literature
Deep web ≠ dark web

What search engines can’t access
What search engines can’t “read”
What search engines won’t index
What search engines can’t or won’t easily surface

Findable with a different mindset!
What’s in the deep web?

Databases

Images, charts, video

Books, articles, documents, reports

Much of social media (more on that later)
What’s grey literature?

Publications not produced by commercial publishers

Universities – dissertations, preprints
Research institutes – clinical trial results, analyses
Think tanks – policy papers, reports

[...]
What’s grey literature?

[...] Governments & NGOs – working papers, data, studies
Associations – reports, newsletters
Conference proceedings
For-profits – white papers
What’s grey literature?

Non-traditional grey lit
  Slidedecks
  Social media, blogs, podcasts
  Patents, standards
  Software
  Survey responses
  Videos
Content from >10 years ago

Google won’t retrieve it
Duckduckgo and millionshort might
If you know the URL, try archive.org
The second page of results...

Organic Result Behavior vs Age

- I normally click on the 1st or 2nd result.
- I scan the first page and click on whatever appears to be the most relevant answer.
- I look for content from websites or brands that I trust and click those results.
- I browse through multiple results on the 1st page.
- I click to the 2nd, 3rd or 4th pages of Google to see as many results as possible.
- I don’t know how I interact with the organic results.

BatesInfo.com N=2129, these are statistically significant differences at a 95% confidence level, not at the 99% level
Why dark/grey content?

More current than published content

Outliers, perspectives not covered in traditional media

Niche focus not well covered elsewhere

Helps fight IAOTWFF
Search Strategies for Grey Literature
Keep thinking like a detective

Look for clues, not the answer

Look for the page BEFORE the dataset

Use your peripheral vision

Be prepared to wade through LOTS of results!
Googling grey lit

Try both Google and scholar.google.de

Use likely words

“technical report” OR “conference proceedings” OR “white paper”

“institutional repository” OR “open access repository”
Finding deep web sources

Look for likely sources:
- Government: statistics, market research
- World Bank, UN, other NGOs
- Associations in your org’s field

Find deep-web monitors:
- infodocket.com, researchbuzz.me
Build your collection

ssrn.com for preprints, other early scholarly research

arXiv.org for hard and social sciences

Researchgate.com?

Copyright issues
Are your org’s work included?
Search strategies

Dig deeper with Millionshort.com
Eliminates most popular sites
Find obscure or less-commercial sites
Search strategies

Know your search criteria:
  Format
  Source
  Timeliness
  Reliability
  Breadth/depth
Search strategies

Start with a libguide

Search with inurl:libguides

Find one good source and look for mentions of it

e.g., “fao.org/faostat/”
Search strategies

ID relevant conferences through general search
Then search conference page for papers, slides, etc.
Try archive.org for prior years
THANK YOU ALL FOR ATTENDING COMPUTERS IN LIBRARIES 2019!

SAVE THE DATE:
COMPUTERS IN LIBRARIES 2020
MARCH 31 - APRIL 2, 2020
HYATT REGENCY CRYSTAL CITY
Computers In Libraries 2019 Presentations

**DISCOVERY, NAVIGATION & SEARCH**

- **A101: Super Searcher Shares Tips!**
  - Mary Ellen Bates - A101_Bates.pdf (1 MB)

- **A102: Search 8.0: Vocal, Graphical, & the Rise of AI**
  - Greg Notess - A102_Notess.pdf (1 MB)

- **A103: Services, Tools, & Techniques for Discovery**

- **A104/A105: Search & the Future**
  - Amy Affelt - A104-A105_Affelt.pptx (2 MB)
  - Tara Calishain - A104-A105_Calishain.pptx (2 MB)
  - Richard P. Hulser - A104-A105_Hulser.pdf (66 MB)
Search strategies

Use fee-based services’ “Source type” to ID grey lit
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-list Blogs</td>
<td></td>
</tr>
<tr>
<td>Business Sources</td>
<td></td>
</tr>
<tr>
<td>Business to Consumer Services (B2C)</td>
<td></td>
</tr>
<tr>
<td>Dow Jones Sources</td>
<td></td>
</tr>
<tr>
<td>Economics and Country Analysis</td>
<td></td>
</tr>
<tr>
<td>European Union Sources</td>
<td></td>
</tr>
<tr>
<td>General Interest Sources</td>
<td></td>
</tr>
<tr>
<td>Government and Politics</td>
<td></td>
</tr>
<tr>
<td>Legal Sources</td>
<td></td>
</tr>
<tr>
<td>Newsletters</td>
<td></td>
</tr>
<tr>
<td>Newspapers: All</td>
<td></td>
</tr>
<tr>
<td>Nongovernmental organizations (NGO)</td>
<td></td>
</tr>
<tr>
<td>Official Government Sources</td>
<td></td>
</tr>
<tr>
<td>Research Reports</td>
<td></td>
</tr>
<tr>
<td>Sports</td>
<td></td>
</tr>
<tr>
<td>Think Tanks</td>
<td></td>
</tr>
<tr>
<td>Top Industry Sources</td>
<td></td>
</tr>
<tr>
<td>Trade (B2B)</td>
<td></td>
</tr>
<tr>
<td>Transcripts: All</td>
<td></td>
</tr>
</tbody>
</table>
Some Grey Lit/Deep Web Tools
Finding theses & dissertations

EBSCO Open Dissertations
biblioboard.com/opendissertations

Open Access Theses & Dissertations
oatd.org

Networked Digital Library of Theses & Dissertations
search.ndltd.org
And follow those leads...

Where has the graduate written, spoken at since then?

Who is s/he collaborating with?
Finding stats and facts

Zanran.com

Uses computer vision to find graphs, stats
Helpful to ID resources to search deeper
Mouseover each result to see graphic
Useful if you don’t know who knows the answer
Data & statistics on sunflower seed exports – 830 results

**Sunflower seed exports (thousand tons)**
www.responsibleagroinvestment.org/rai/sites/responsibleagroinve
seed exports (thousand tons) Sunflower seed Marketing year (S
1.4 Sunflower seed sale prices As with many other commodity m
of sunflower seed prices in the spring of 2008 due to a poor sun
seed prices started
2102 (projection) | INSIDE RAI | RAI Knowledge Exchange Pla

**Ukraine commitment Year (Agreement) Ukraine export duty t
broken as covered below**
Year (WTO) Ukraine commitment Year (Agreement) Ukraine export duty to EU Trigger Level (tonne) Maximum surcharge For the sunflower seeds, whether broken as covered below: WTO E1F+11 E1F+12 E1F+13 E1F+14 E1F+15 EN EN ...

**12030000 Copra 12040000 Linseed 12051010 Rape or colza seeds of low erucic acid 12051090 Other low erucic acid rape or colza seeds 120590 or colza seeds. nes 12059090 Rape or colza. nes...**
gs.mofcom.gov.cn/accessory/200804/1208158953579.pdf
or colza, nes 12060010 Sunflower seeds for sowing 12060090 Sunflower seeds excl...
Inulin 11090000 Wheat gluten 12010010 Soya beans seed 12010091 Yellow soya beans excl. seed 12010092 Black soya beans excl. seed 12010093 Gr beans excl. seed
Jan 2019 | ?????????????????? – 1 more result from this site URL
Finding datasets & statistics

Use search engines for leads:

Keywords (database OR dataset)
Keywords (filetype:xls OR filetype:xlsx)
Finding datasets & statistics

Datacite: open-access datasets, articles, images, etc.
Only searching metadata
Good for granular searching
search.datacite.org
Finding datasets & statistics

Registry of Research Data Repositories
Datacite’s directory of data collections
Only searching metadata
Good for searching broadly
re3data.org
Finding deep web content

Start with one known source
Association, NGO, non-profit, library or museum, etc.

Google: *keywords* (portal OR resources OR “online tool”)

BatesInfo.com
Curate open access resources

Open-access journals: doaj.org

Open-access books: doabooks.org

Open-access repositories: opendoar.org
Curate open access resources

Public Library of Science: plos.org

OpenGrey (Grey Lit in Europe)  
opengrey.eu

BASE – Bielefeld Academic Search  
Engine  
base-search.net
Questions?