



# Finding Hidden Web Resources

**Mary Ellen Bates**

**BatesInfo.com**

**March 12, 2020**



# What we'll cover

**What's deep web content? What's grey literature?**

**Search strategies for deep web/grey lit**

**Build your own deep web collection**



# Deep web ≠ dark web

## Deep web is:

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**

**Dynamic content**



# What's grey literature?

**Publications not produced by commercial publishers**

**University dissertations, preprints**

**Research institute reports**

**Think tank policy papers, reports**

**Government & NGO working papers, data, studies**

**Association reports, newsletters**

**Conference proceedings**

**[...]**



# What's grey literature?

**For-profit white papers, manuals**

**Slidedecks**

**Social media, blogs, podcasts, comments**

**Patents, standards**

**Survey responses**

**Videos, audio**

**Catalogs**

**etc...**



# What's grey literature?

**Content from >10 years ago**

**Google won't retrieve it**

**Duckduckgo and millionshort might**

**Think creatively – [archive.org](https://archive.org), fee-based online**



# Clients won't ask for grey lit

**Your clients only ask you to do what they *think* you  
*can do***

**Use the ref interview to expand their vision of an  
info pro**





# Search Strategies for Grey Literature



# Think like a detective

**Look for clues, not the answer**

**Look for the page BEFORE the resource**

**Use your peripheral vision**

**Be prepared to wade through LOTS of results!**



# Search strategies

**Know your search criteria and filters:**

**Format**

**Source**

**Timeliness**

**Reliability**

**Breadth/depth**



# Search strategies

**Start with a libguide**

Search with **`inurl:libguides`**

**Find one good source and search for mentions of it**

e.g., **`"afdc.energy.gov"`**



# Googling grey lit

Try both Google and Google Scholar

Use likely words

"technical report" OR "conference proceedings"  
OR "white paper"

"institutional repository" OR "open access  
repository"



# Googling grey lit

## Use likely filters

`filetype:pdf OR filetype:doc OR filetype:docx`

`site:.edu OR site:.gov`



# Use site: to expand

**Expand to dig deeper in a resource**

**Find pages suppressed or not indexed by site**

**Difference relevance ranking = different results**



# Use site: to explore

Find *subdomains* to explore within a web site

`site:nasa.gov -site:www.nasa.gov`

Shows other areas of web site

`astrobiology.nasa.gov, eyes.nasa.gov, exoplanets.nasa.gov ...`





# Go beyond Google

**Dig deeper with Millionshort.com**

**Eliminates most popular sites**

**Find obscure or less-commercial sites**

**ID a site with instant responses**

## FILTERS

### POPULARITY

- Don't remove any
- Remove Top 100
- Remove Top 1000
- Remove Top 10k
- Remove Top 100k
- Remove Top Million

### E-COMMERCE

- Don't remove any
- Show ecommerce sites
- Remove ecommerce sites

### LIVE CHAT

- Show all sites
- Show sites with chat
- Remove sites with chat



# Going beyond the web

**Use fee-based services' "Source type" to ID grey lit**  
**Not comprehensive but useful**



Source type: °

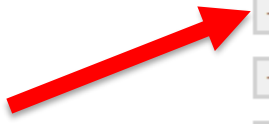
- Select all
- Audio & Video Works
- Blogs, Podcasts, & Websites
- Books
- Conference Papers & Proceedings
- Dissertations & Theses
- Encyclopedias & Reference Works
- Government & Official Publications
- Magazines
- Newspapers
- Other Sources
- Pamphlets & Ephemeral Works
- Reports
- Scholarly Journals
- Standards & Practice Guidelines
- Trade Journals
- Wire Feeds
- Working Papers





Select Source Category By Type ▾

- + A-list Blogs *i*
- + Business Sources *i*
- + Business to Consumer Services (B2C) *i*
- + Dow Jones Sources *i*
- + Economics and Country Analysis *i*
- + European Union Sources *i*
- + General Interest Sources *i*
- + Government and Politics *i*
- + Legal Sources *i*
- + Newsletters *i*
- + Newspapers: All *i*
- + Nongovernmental organizations (NGO) *i*
- + Official Government Sources *i*
- + Research Reports *i*
- + Sports *i*
- + Think Tanks *i*
- + Top Industry Sources *i*
- + Trade (B2B) *i*
- + Transcripts: All *i*





# Build your own Grey Lit / Deep Web collection



# **Build a collection for *your* clients**

**Look for likely sources:**

**Government: statistics, market research**

**Key associations and NGOs**

**Find deep-web monitors:**

**[infodocket.com](http://infodocket.com), [researchbuzz.me](http://researchbuzz.me)**

**See Jennifer Klang's 2019 slide deck for other sources**



# Finding datasets & statistics

**Datacite: open-access datasets, articles, images, etc.**

**Only searching metadata, limited search functions**

**Good for granular searching**

**[search.datacite.org](https://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](https://re3data.org)**





# Curate open access resources

**Open-access journals: [doaj.org](http://doaj.org)**

**Open-access books: [doabooks.org](http://doabooks.org)**

**Open-access repositories: [opendoar.org](http://opendoar.org)**



# Talk to people!

**Find a librarian to talk to**

**LinkedIn advanced people search**

**Industry = Libraries**      (yeah, go figure)

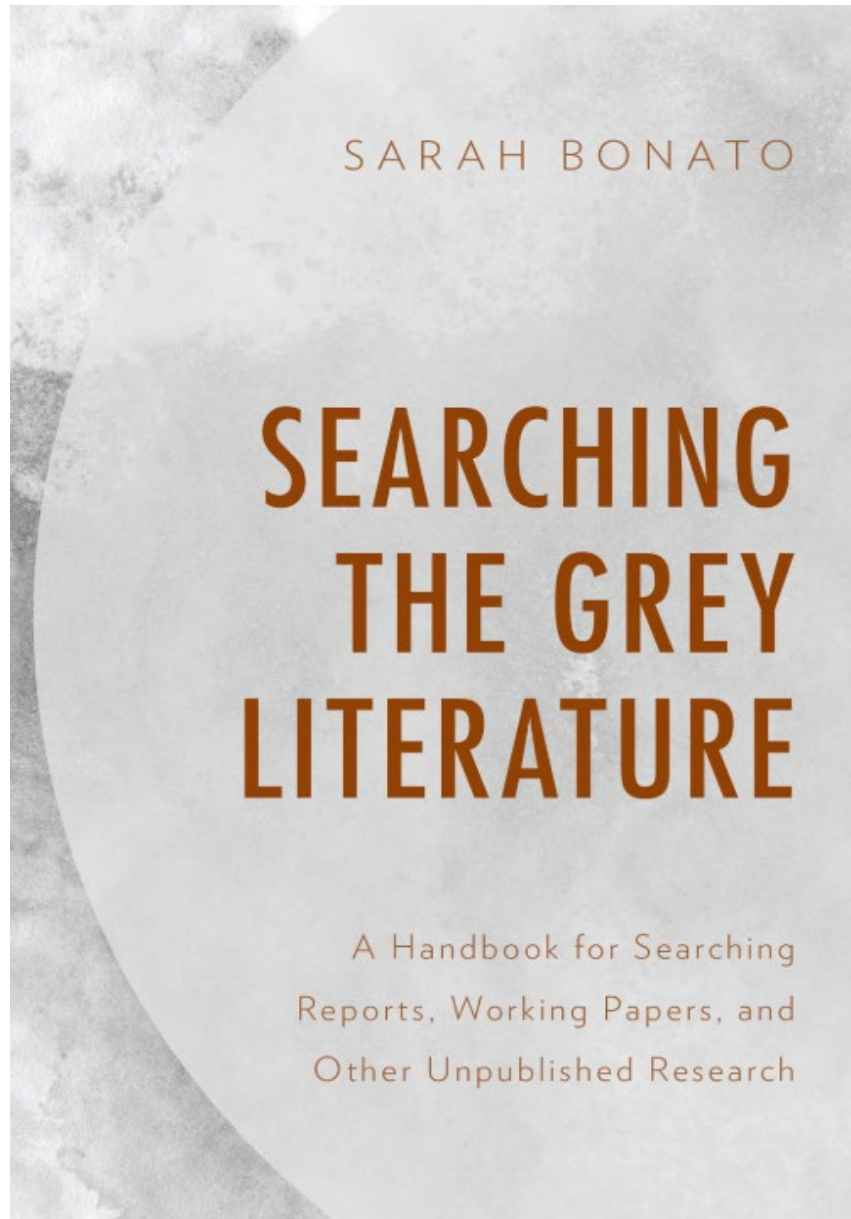


# Final thoughts...

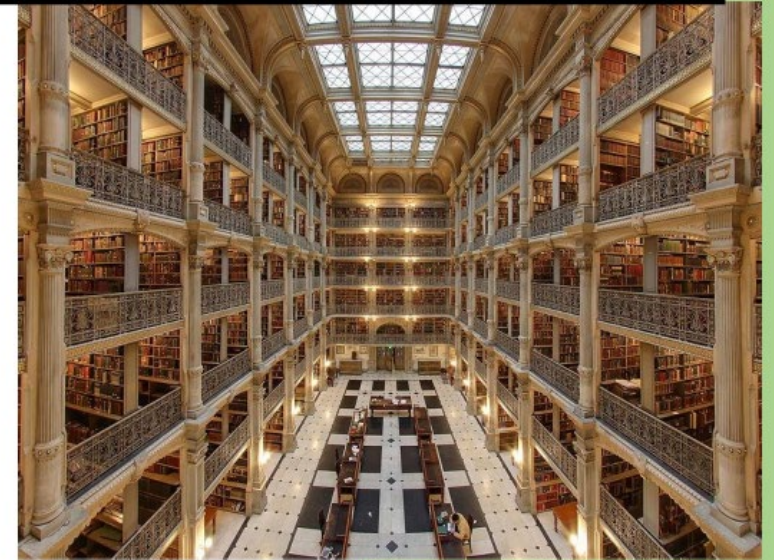
**Be creative – grey literature is hiding in plain sight**

**Clients won't ask for grey literature – look anyway**

**No grey lit search is comprehensive – go Zen**



## Best of the Best of the Business Web



Robert Berkman, Editor



# Let's stay in touch!

**Mary Ellen Bates**

**mbates@BatesInfo.com**

**BatesInfo.com**

**+1 303 772 7095**

**LinkedIn: maryellenbates**

**Facebook: maryellenbates**