Discovery Systems? Or Gov Docs Usage Decreasing Systems?

Kathryn C. Millis
DePauw University Libraries
0172

Millis - Discovery Systems - FDLP 2014

1

Welcome.

This is the presentation on "Discovery" systems.

I'm Kathryn Millis, government information (and other stuff) librarian at DePauw University Libraries.

We're depository #0172.

"Web scale" "discovery" systems

- Increase access to government documents?
- Hide them, decreasing, perhaps even decimating, your patrons ability to find and use government information?
- Ex Libris's Primo Ebsco's EDS WorldCat
 - vs. catalog.gpo.gov & my opac
- · Not promote, endorse, or recommend
- Gov doc librarians can ...

Millis - Discovery Systems - FDLP 2014

2

The abstract in our program says I'm talking about "web scale" "discovery" systems, and whether they might increase access to official U.S. government information, or might instead "hide them, decreasing, perhaps even decimating your patrons ability to find and use government information."

Libraries are rapidly moving to "discovery" systems. Many libraries promote and feature them prominently on our home pages, while the catalog and databases are clicks away, or they may even replace our catalogs. It's important that these systems help people find government documents on topics of government interest, about which the government has published important, often primary, resources.

I said I'd talk using Ex Libris's Primo, Ebsco's EDS & WorldCat's WMS, comparing them to the Catalog of Government Publications (or CGP) and my library's catalog. If you're specifically interested in Ebsco, I apologize, but I won't address them. I realized that I can't make an fair comparison. My catalog, Primo, and WorldCat all search things my library provides, but I had to use a different library's collections and access when searching Ebsco.

As promised, I'm not promoting, endorsing, or recommending any specific discovery layer, or how they as a whole help our researchers to find government information. I do have a few suggestions of things we can try to increase access to government information online and in print, for our patrons when they use our "discovery" tools.

Today

- Discovery
- · What people pick / click
- · Results from simple, sample searches
- · What we can do

Millis - Discovery Systems - FDLP 2014

This presentation has four basic parts.

First, I need to give a quick overview of what "discovery" systems have and do, as I understand them.

Second, we'll consider one useful model of how people pick what they click from a search tool's result list.

Third, the main part, we'll look a the results from some simple, sample searches I did in the CGP, my opac, Primo, and WorldCat.

Fourth, I'll suggest a few things we librarians can do or try to do.

1. Discovery

- "web scale" "like google" big database of:
 - Your catalog
 - Bibliographic records! Not full text!
 - Full text:
 - Gale OneFile or Ebsco Academic Search Complete or ...
 - Bibliographic with or without subjects & abstracts
 - PsycINFO, Early American Imprints, ...
 - Really different
 - Web of Science (citation index)
 - Video, music score databases

Millis - Discovery Systems - FDLP 2014

-4

Part One.

"Discovery tools" or "discovery layers" are the latest search tools; vendors & librarians describe them as "web scale" or "like Google."

"Web scale" is fairly safe, because people will either think it means "on the internet" or "in the cloud, so it's not actually on my computer but I can access it" or something.

"Like Google" is very dangerous, because it suggests things that are not true.

First, some degree of comprehensiveness. Google searches (people probably mostly think) darned near the whole web or close enough for now, billions and billions of things. Discovery layers search only a subset of your library's resources.

Second, Google searches the full text of web sites (& books, magazines & patents). In many cases (for instance, our catalogs) discovery layers can't search items' full text.

Third, and worst, Google sorts results very well putting relevant materials at the top of the result list, so saying they're "like Google" suggests that discovery layers do too.



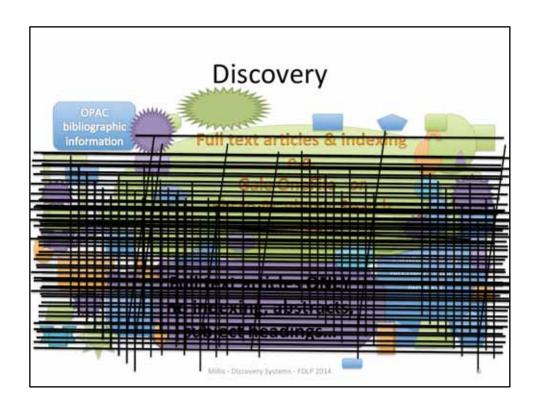
Basically, a discovery layer makes a big pot from a bunch of information sources & searches them simultaneously.

It generally includes, and visually, we might simplify it to these types of ingredients:

- bibliographic info only (from your catalog, other catalogs, and similar sources)
- full text articles with bibliographic and descriptive info (from one or two very big, and many, many smaller sources)
- full text articles with some or none of that (from a bunch of sources)
- specialized databases that it really can't do what the native interface does, like citation databases, audio or video recordings, statistics, or musical scores
- etc. etc. Lots of etc.

And then they get processed and indexed so the "discovery" tool can do things like an Author search through author names in all of them, or a title search.

But many of these databases and contents are resources that are available only by subscription. Your library doesn't subscribe to them all, and it doesn't give you access to things you don't subscribe to.



Instead, it runs the results through a sieve or filter and shows only what you subscribe to or can get free.

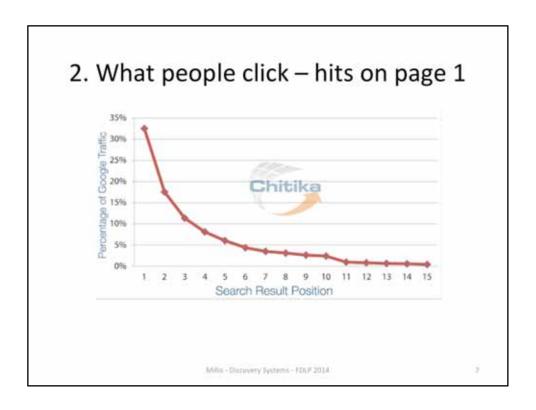
Or you can set it to show records for things you don't have, so your patrons can make InterLibrary Loan requests.

But getting a discovery layer doesn't give you or your patrons anything you don't already have access to via subscription or freely.

It also is extremely unlikely to include most of the databases you subscribe to. But I digress.

So that's part One, "discovery layers."

Anybody want to suggest a revision to this explanation? Did I say anything that seems untrue or unfairly misleading?



Part Two.

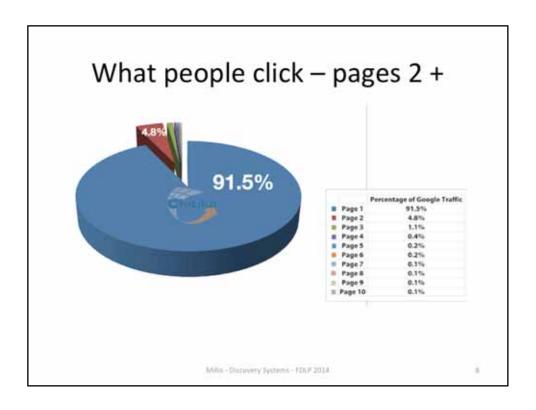
Even if librarians don't make any comparisons to Google, it is so heavily and widely used that it influences what our patrons (the public, students, faculty, & staff) and we ourselves expect from a search tool. We all know what Google can do.

Google has gotten really good at searching and sorting it so that users expect to be satisfied by top hits. This image, from a widely reported study, shows how much people use the first few results compared to those just a little lower down on the list. According to this analysis, almost 35% of our clicks go to the top item on Google's results. As we scroll down, we rapidly get less likely to click something. By the eleventh item, we're down to only about 1%! (That's still millions or billions of clicks.)

People do this successfully, many times a day. We often are, & increasingly we expect to be, satisfied by the top results. We trust Google, & by extension, what it shows us.

But when searchers turn to our library catalogs and databases, they routinely have to go past the first two or three (or twenty or thirty or more) results to find what they really want, even though vendors & we claim that they too sort by "relevance."

One interesting note: this chart is about the first page of results, but if people go to pages 2, 3, etc., Chitika says we're again far more likely to click the first hit on that



We are, however, not likely to go to page two. Over 91% of peoples' clicks are to results on page one. If we click through to the second page of results we are unlikely to follow any of those links. Fewer than 5% of clicks are on page two.

And we're even less likely to follow links on the third, fourth, or subsequent pages.

People are pretty sure Google's first results are best. (And there's reason to be concerned about our increasing trust, but that's a different project.)

The defaults for the tools I used are ten hits per page.

If that's "true"

- Gov Docs need to be high on results lists or people won't "discover" them
- Truer if we call it "one" or "comprehensive"?
- Less true when?
 - · Obvious date sort.
 - · Obvious clusters.

Millis - Discovery Systems - FDLF 2014

10-

To the degree that people have similar expectations (or at least hopes) of our library search tools, they (and I'm saying "they" because librarians presumably have knowledge & experience that give us different expectations from those of the general public) will expect to pick from the top handful of results. They will expect the first to be the "best," and will not normally scroll down past things that look irrelevant or unhelpful in the hopes that things farther down will meet their needs.

So if government documents are particularly appropriate for a topic, people need to see them high on the result list.

I think this is more likely when we use names or language that suggests breadth, depth, and comprehensiveness, calling a discovery layer "One Search" or saying it "searches our collections," or "our resources," or even "millions of things." (Yes, they are broad and deep, but not comprehensive!) Language like this sets and encourages high expectations.

I think this is less true when there are obvious clues that it's neither comprehensive nor sorting by relevance. Saying it searches "some databases," or obviously having a default sort by date don't suggest to people that the first two or three results will meet their needs better than the results after them (or not there at all).

3. Searches

- catalog.GPO.gov
 - · Ex Libris
- opac
 - · Ex Libris
 - Semi-shared with PALNI (24 colleges & seminaries)
- Primo
 - Ex Libris
 - Semi-shared with PALNI
- WorldCat local
 - Not Live
 - Shared with PALNI
 - default search "libraries worldwide"

Millis - Discovery Systems - FDLP 2014

10

Part three.

I wanted to know if discovery layers would put government documents high enough in results that searchers would, if Chitika's study applies, be likely to pick them.

For comparison, I first searched the Catalog of Government Publications (CGP), to see how many documents it found. CGP only approaches comprehensiveness for things published since 1976, though it has some older records. I picked recent topics so that we could have a vaguely, approximately, accurate estimate of results.

I next searched my library's catalog, which comes from Ex Libris, the same company that does the CGP. DePauw is in a consortium, called PALNI, of colleges and seminaries, but our catalog shows only our holdings.

Then I searched two discovery layers, Primo & WorldCat. DePauw & PALNI have had Primo, an Ex Libris product, for several years; each school's patrons see only their own holdings & subscriptions.

At the end of June, PALNI will switch to WorldCat. "WMS" is still in testing, so I did these searches using WorldCat Local. PALNI's board has decided several things that heavily impact results. For one, the default search is no longer DePauw's holdings, but "libraries worldwide." A possible upside here is that my patrons might learn about government documents we haven't cataloged!

My results impacted by:

- Started routine cataloging 1996
- · Dropped & weeded print
- No gov doc database subscriptions
- No electronic only records in opac

Millis - Discovery Systems - FDLF 2014

11.

Several things have big impacts on my patrons' ability to find government information when they use DePauw's catalog or discovery layers.

First, we didn't start routinely cataloging government publications until 1996. Before then, we put a few things in the general or reference collections, but the vast majority were left in gov docs and out of the catalog.

Second, I've dropped and weeded our print holdings. We no longer get hearings, reports, etc., and I've withdrawn hundreds of shelves, much of which we never cataloged, and I couldn't make the case that we should. So my students and faculty and the general public aren't going to find a dozen hearings (or several dozen) on Daylight Saving Time, NPR funding, acid rain, child porn, the Clinton administration's goal of national health care, the ADA, Title IX, POWs, etc. when they search our catalog or discovery tools for those topics.

We don't subscribe to Marcive's "Documents Without Shelves," don't subscribe to the GPO database via Ebsco, FirstSearch, or ProQuest. And we don't load opac records based on new electronic titles lists (or any others).

Any of these could add (at a cost) access to at least electronic documents.

Discover? Or Hide?

- · How many total?
- · How high is first easily identifiable gov doc?
- · What's above them?

Millis - Discovery Systems - FDUF 2011

12

But we do have government documents, cataloged, both print and links to online versions of documents we hold in print. Many of them are valuable primary sources from appropriate agencies, which my library's search tools should make readily available to people interested in the topics they cover.

I've used these tools a lot, and for this project, specifically looked at:

How many government documents did I find when I searched?

How high on the list is the first easily identifiable government document? High enough that (to the degree that the Chikita study is applicable) people are likely to pick it?

When results are sorted by the default, which is "relevance," what's above them? What do Primo and WorldCat say is more "relevant" than them?

My searches

- 9/11
- Biofuels
- Death tax
- · Don't ask don't tell

Millis - Discovery Systems - FDLP 2014

11

I chose four basic topics that have been covered significantly and authoritatively by federal government publications.

They're neither brand new (too new for our library to have received print documents about them) nor historic (too old for us to have cataloged documents about them when we originally received them).

The U.S. government has published many, not just a few, things about these topics during the last ten or fifteen years.

They're clearly "government" related topics. Unlike literature or the arts, or even sciences, average citizens know these topics are government topics.



My first search in each case was in the Catalog of Government Publications, or "CGP" at catalog.gpo.gov.

It failed when I typed nine slash eleven, finding everything with the number 9 and also the number 11, anywhere in the record. So I had to put "9/11" in quote marks. (That's the only time I had to do anything "advanced." In all other searches I just typed nine slash eleven, or don't ask don't tell with no quotes or caps or anything.)

When I searched "9/11" I got 1,423 results. Because there are unique records for each format (paper, microform, and online), there really aren't 1,423 different documents. Here we can see that one and two are the same thing in two formats; five, six, and seven are the same thing. So rather than 1,423, let's guess, just approximately, probably more like 1,000 unique titles.

They're clearly arranged in chronological order. Results in a table, with the year smack in the center column, make that pretty easy to figure out. It might not be the first thing an average person will notice, but it's reasonably easy to tell that these are the newest, with no claim to being the most "relevant" or "best" results.

Personally, I don't believe it's possible, with bibliographic data only, to rank by relevancy in any but the most general way. Probably all of these have "9/11" in the title, and little else that could help sort them beyond that.



In my library's catalog, like yours and the CGP, we're looking only through bibliographic information, no full text, and we're finding records for books, CDs, maps, periodical subscriptions, and so on, including but not only government documents.

It's not as obvious as with CGP, but results are in chronological order, with the most recently acquired (not most recently published) at the top. Again, these probably have "9/11" in the title. Some may also have it in the table of contents. But we have only that one term, probably only once, maybe twice or three times, while Google also has full text, and over a decade of surveillance tracking things like which hits people click to help them figure out what people want when they Google 9/11.

The first two items on the list are books about Bush & Cheney, and about spying in the NYPD.

The first government documents my search for nine slash eleven found were items numbered three and four, so still pretty high up. Desktop users barely need to scroll down to see them, though laptop users definitely do. (Our students are required to have laptops, so I'm pretty sure this has an impact.)

Sadly, they don't have cover art available to attract potential readers.

First 3 gov docs via opac

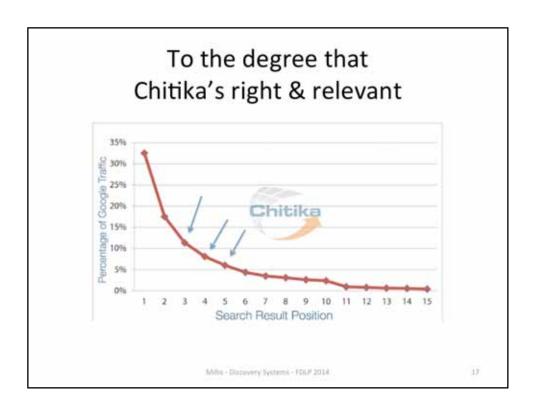
- #3 "Eleven years after 9/11: can TSA evolve to meet the next terrorist threat?: hearing..."
- #4 "A decade after 9/11 could American flight schools still unknowingly be training terrorists?: hearing..."
- #5 "Terrorist financing since 9/11: assessing an evolving al-Qaeda and state sponsors of terrorism: hearing..."

Millis - Discovery Systems - FELF 2014

16

So in my library's catalog, the first three government documents my search for 9/11 found are congressional hearings, with titles starting "Eleven years after 9/11," which is the third item on my result list; "A decade after 9/11," with is the fourth item on my list, and "Terrorist financing since 9/11," which is the fifth item on my list.

Co-incidentally, their titles help us realize that our first results are recent.



So, if the Chitika study can be applied to result lists from library catalogs and other library search tools that sort things by date, we might imagine that *theoretically*, *approximately* 35% of users' clicks might go to the Bush Cheney book, 18% go to the NYPD book, and for the first three gov docs I found, 10% of clicks would be to the hearings on ten and eleven years after 9/11, and 5% would go to the hearing on financing.

Again, it's not a perfect comparison. But worth considering. Any objections or concerns about how this seems to relate?

I'm not suggesting it's exact.

Or even close.

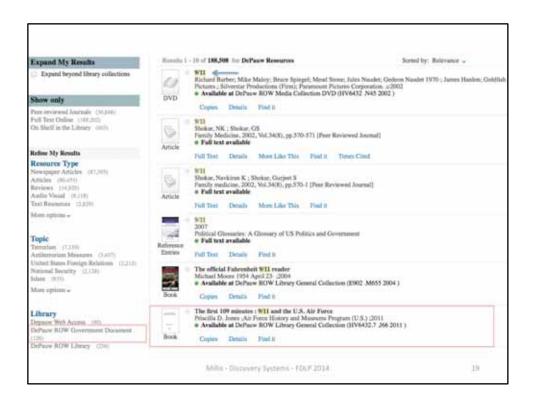
Millis - Discovery Systems - FDLF 2014

18

I don't think that's true, or close, but it's really worth thinking about how much our library tools require our patrons to behave in ways that are increasingly unnatural (to them) to get to official government information (in the form of government documents).

Chronology is not the main problem!

Also, I assume that most people would look into several, perhaps "many" items they find. It's not like 10% of people would go to the book on eleven years after 9/11 and no other book. Nor do people using Google pick one and only one item from their result list.



Third, I searched 9/11 in Primo, our discovery layer from Ex Libris. So now we're looking through our catalog and a bunch of databases. We have it set to prioritize catalog records, so other things being equal (in the bibliographic information and Ex Libris' algorithms), a gov doc or other book should appear above (or more "relevant" than) an article with similar characteristics.

Let's notice a few things. First, upper right, it claims to sort by relevance. Second, of the 188,508 things it found, it claims that the four most relevant have single word titles which exactly match my search term, "9/11."

In the left column we have ways to limit those 188,508 things. But we're supposed to say "facet" now. We can "facet" to get only articles, or only things on terrorism (7,000 things out of 188,508, because different databases use different terminology & it's not merged or massaged or whatever we're supposed to call that) or things that are in DePauw's "ROW Government Document" collection. One possible upside is that this tells people there is such a place and it contains material on 9/11.

On the other hand, if gov docs aren't high on the list, is that it tells them that other things are better and more relevant than government documents.

The first gov doc is hit number six, but notice that like Google, they're not numbered.

First 3 gov docs via Primo

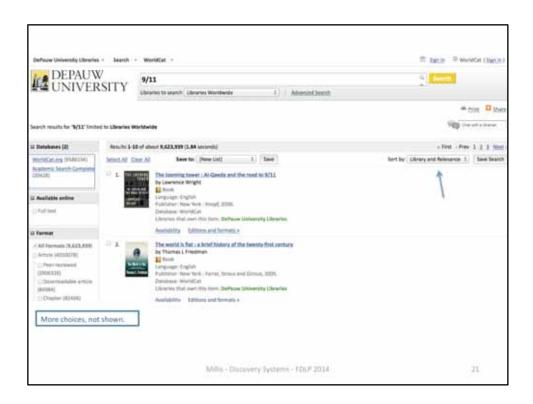
- 1. #6 "The first 109 minutes: 9/11 and the U.S. Air Force"
- 2. #15 "Pentagon 9/11"
- # 35 "Attack on the Pentagon : the medical response to 9/11"

Millis - Discovery Systems - FDLP 2014

30

So, the first government document on my Primo result list, is hit number 6, about 9/11 and the Air Force. The second document is on page 2, hit #15. And the third is on page four, hit #35, on the medical response to 9/11.

Chitika had only about 5% of clicks occurring on page 2 of results, and 0.4% on page 4 of the results. For people to make it to "Attack on the Pentagon," they must be acting very abnormally!



My fourth & final search for 9/11 was in WorldCat Local for DePauw, which we are still working on. It isn't live yet, and is still "only" searching WorldCat and one database, Academic Search Complete. But it still found over 9 million things, because it's set to search "Libraries Worldwide."

On the middle right, you see its default sort is by two factors: relevance & whether or not DePauw has it. Stuff DePauw has ranks higher than stuff we don't. After DePauw, stuff any PALNI library has ranks higher than stuff nobody in PALNI has. This has a terrible impact on access to gov docs. Few PALNI libraries are depositories, so after stuff DePauw has, WorldCat next shows searchers the non-gov docs held by PALNI libraries. And only after that does it show them gov docs that are held by hundreds, thousands, or tens of thousands of other libraries.

In the left column, again, we can limit or "facet" by format, language, author, date, etc. If you scroll all the way to the bottom of the screen (not shown) past about 100 possible choices, one possible "topic" is "government documents." That's better than nothing, but even if people don't imagine the facets are also in order of relevance, it's still as far down as you can go.

Anyway, the default result is that the most "relevant" thing we have is a book by Lawrence Wright, "The looming tower," and the second most "relevant" thing is Friedman's book "The world is flat."

First 3 gov docs via WorldCat

- #9 "The 9/11 Commission report : final report of the National Commission on Terrorist Attacks upon the United States"
- #20 "Our White House: looking in, looking out" (kids book)
- #50 "Commemorative joint meeting of the Congress of the United States in remembrance of the victims and heroes of September 11, 2001, Federal Hall, New York, NY, Friday, September 6, 2002"

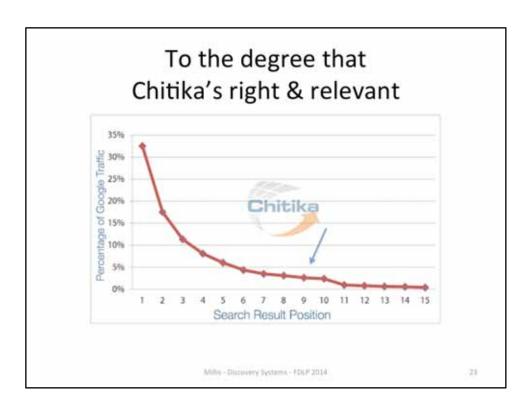
Millis - Discovery Systems - FDLF 2014

22

The first gov doc this search found is hit #9, the final report from the official, national Commission. That's an important, official, primary source. I'd have ranked it above Friedman's book! Wouldn't any of us?

To find the second government document, searchers would next have to go to the second page of results. And scroll as far as they can, seeing and presumably rejecting or at least not being satisfied by every other thing on the list as they go. The second gov doc is a children's book about the White House, #20, at the bottom of page 2.

And the third gov doc is hit #50! At the bottom of page 5, where on a Google search we'd be at about 0.2% of clicks.



So to the degree that we can apply Chitika to our searches, probably that official final report would get clicked less than 5% of the time. And the other gov docs will likely be unseen and unclicked.

	Total	1 st doc	2 nd	3rd	Also docs:
GPO	13,850	1	2	3	all
"9/11"	1,423	1	2	3	all
opac	335	3	4	5	6, 7, 8, 23, 24, 25, 26, 27, 28, 42, 43, 45, 52
Primo	188,508	6	15	35	
WorldCat	9,630282	9	20	50	

Here's a summary of how these four tools did at displaying government documents about 9/11 near the top of their result lists when I searched for 9/11.

GPO's catalog is all gov docs, so the top results were government documents. As were all the others.

My library's catalog got government documents as high as the third item, because it's chronological & we haven't yet purchased newer things with 9/11 in the title or table of contents. (Sooner or later, we will.) When I kept scrolling down through about the first fifty, government documents were also items number 6, 7, 8, 23, 24, 25, 26, 27, 28, 42, 43, 45, and 52.

But when we move to discovery tools, the top, most "relevant" government documents are farther down. With the discovery tool my library currently has, Primo, the first gov doc is #6, and even further down, #9, with the one we're going to have, WorldCat.

In both cases, these three gov docs were the only gov docs in top fifty hits.

What was more relevant?

- Thomas Friedman's "The World is Flat"
- Michael Moore's "Fahrenheit 9/11"
- · "Family Medicine 1 page articles
- · "9/11" entry in reference books

Mills - Discovery Systems - FDLF 2014

25

I don't like that.

Even more troubling, while people trust that Google will show them highly relevant things, let's review what Primo and WorldCat say is more relevant than the government documents!

Assuming we just believe them like we trust Google, government documents must really not be very relevant!

Or maybe we (librarians and our patrons) recognize that these search tools are just not as good as Google and we can't trust them the same way, and must work harder when using them.

Maybe this is temporary. Or a glitch. Or I am too demanding. Who wants government documents about 9/11 anyway?

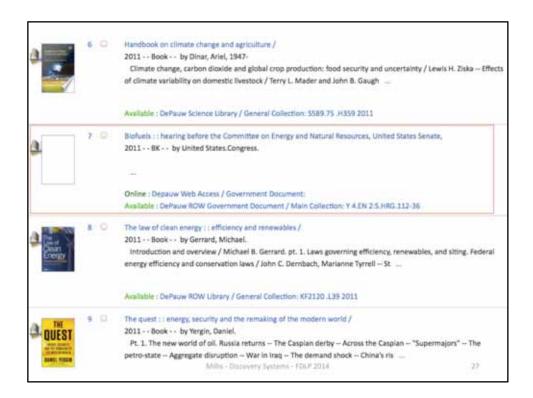


Let's look at my second set of searches, for the single word "biofuels."

Again, the catalog of government publications shows 190 things, but some are duplicates.

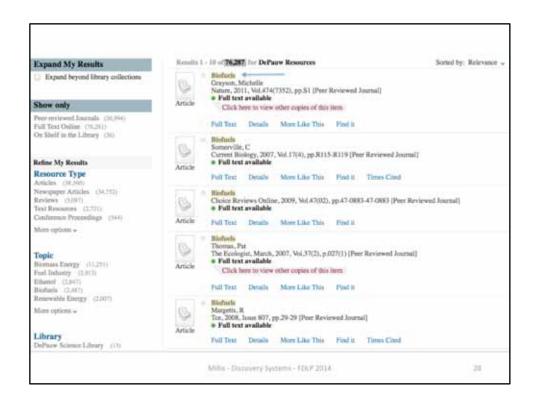
Please note on the far right column, most of these things are online, so my patrons' access to them, and your patrons' access, is not limited by what we physically hold or have cataloged. They are available. The public, and my students, faculty, and staff, ought to find them using our tools, ought not have to go to catalog.gpo.gov to look for them.

They ought to be listed in our discovery tools, without us paying extra for them, just as we don't pay extra to search HathiTrust or other free databases that are included in these tools. (And only once they're listed can we really focus on my question of how they're ranked.)



In my catalog, the first gov doc is the seventh item on the list.

Remember this is chronological by acquisition, so one thing this means is that we have gotten six things with the word "biofuels" in their bibliographic records since we cataloged that. We are apparently interested in biofuels.



In Primo, we again see how ranking appears to be heavily influenced by the length of titles. It appears that a one word title which exactly matches the word I've searched will be ranked high. It's also probably, I think, that the word "biofuels" shows up a lot in the full text of these articles.

Our first gov docs are farther down the list.



WorldCat did better, getting government documents as the third, fourth, and fifth titles on my list. It appears that WorldCat places less weight on the length of the title, because as we know and can see, some gov docs, especially hearings, have very long titles!

	Total	1st doc	2 nd	3rd	Also docs:
GPO	190	1	2	3	all
орас	34	7	10	13	
Primo	76,287	16	37	38	
WorldCat	36,731	3	4	5	6, 7, 8, 9, 10, 11, 13, 14, 18, 19, 23, 25, 36

Here's our summary of how these four tools ranked government documents in my "biofuels" search.

Remember that the first two, GPO and my opac, are chronological. It's only Primo and WorldCat that make claims of "relevance." Primo considers gov docs not very relevant. Only three of the top fifty are gov docs. The "most relevant" doc, hit #16, is on page 2, in the group that would get fewer than 1% of clicks.

WorldCat gets this one better! Eight of the "top" ten results are government documents. And there are a total of sixteen in the top fifty.

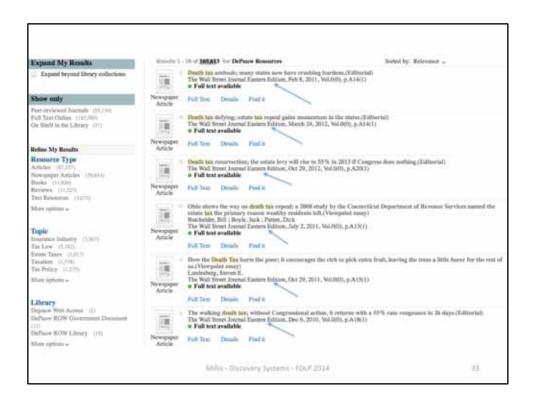


GPO's catalog finds 29 things that use the term death tax, which is a somewhat pejorative, but widely used, term that means "inheritance and transfer tax."

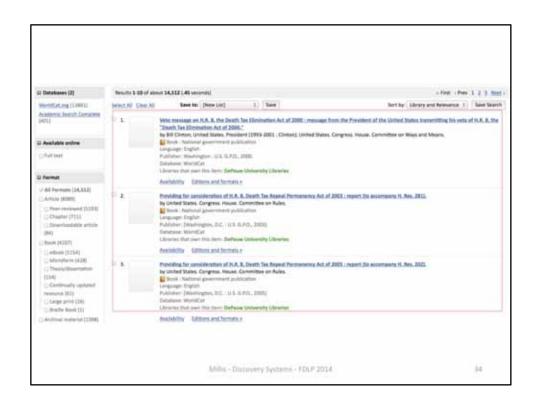
Almost all are online, so if a discovery layer included CGP, our patrons would find and be able to access them.



In my catalog, the first government document with both the word death and also the word tax in the bibliographic record is number ten. Books like number nine, "Art law conversations" also have both words but are more recently acquired.



According to Primo, (at least) the six most relevant things that my library has are opinion pieces from the Wall Street Journal!



WorldCat again put government documents about a government topic at the top of the result list. Each one has "death tax" in the title. Yay!

But I don't know if this will continue once it is searching dozens of databases (like Primo does), not only Academic Search Complete.

	Total	1 st doc	2 nd	3rd	Also docs:
GPO	29	1	2	3	all
opac	29	10	13	14	
Primo	165,613	12	13	13	14, 15, 16, 17, 18, 19, 20, 31, 32
WorldCat	15,512	1	2	3	4, 5, 6, 7, 8, 9, 10, 11, 16

Our catalog's government documents with the words "death" and "tax" barely made it to the first screen of results, because we have bought newer things.

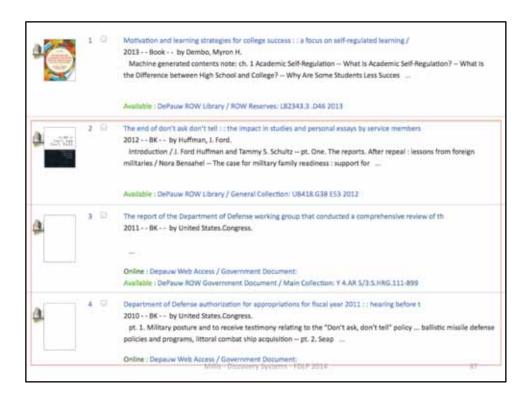
Primo's results are almost as high, but because it claims to rank by relevancy, I expect that my students might think less of them and select them less often.

WorldCat got government documents about the death tax to the very top, and both it and Primo offered my patrons lots of government documents in the top fifty picks, which is good.

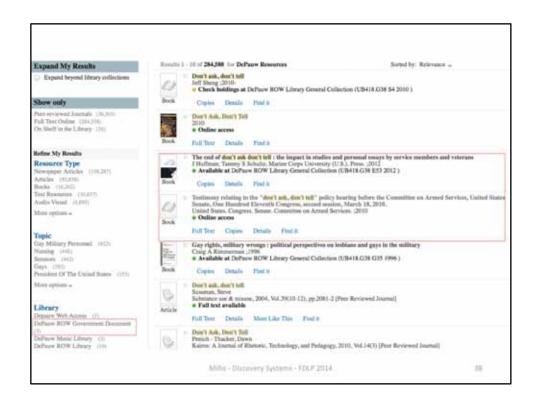


My final search was for don't ask don't tell, which I did not put in quote marks because the average person does not.

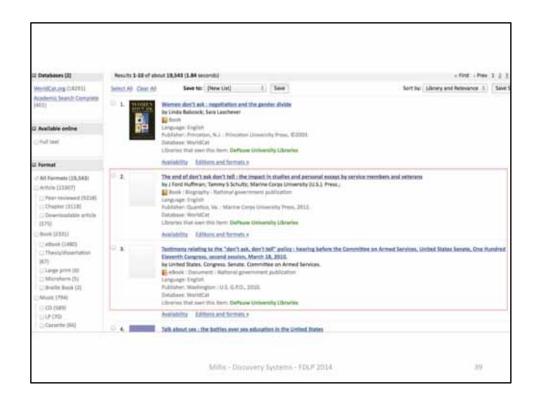
I was really surprised at how few things GPO found! Only twenty five? Or maybe about a dozen, once we account for multiple formats? Something is wrong. But I'm going to accept what it says.



Had I done this search a few years ago, during the gap between passing and repealing "Don't Ask Don't Tell," my catalog might not have had gov doc hits at two, three, and four.



Primo again showed a strong preference for exact title matches, but almost inexplicably slipped in a couple of government documents and a book as results three through five, despite their significantly longer titles.



WorldCat got the exact same gov docs as high as hits two and three.

	Total	1st doc		3rd	Also docs:	
GPO	25	3	2	3 5	all	
opac Primo	284,588		4			
WorldCat		2				

All of the tools did pretty well at getting government documents about Don't Ask Don't Tell to the top of the result lists. But notice that both Primo and WorldCat, while they each had two at the top, also had only one other document in their top fifty hits.

So, does this tell the general public, and my students (and faculty, but I'm more concerned about the public and students, on the theory that faculty members have learned to be a bit more discerning), that the U.S. federal government has published a couple of things, but not all that much that's relevant to this federal policy?

Or is it really helpful to have at least a couple of gov docs show up, as a clue that more may be available someplace?

Titles between #3 & #42

- Andrew Lloyd Webber anthology
- Steal like an artist: 10 things nobody told you about being creative
- Next-day job interview: prepare tonight and get the job tomorrow
- America's food: what you don't know about what you eat
- If the universe is teeming with aliens ... where is everybody?: fifty solutions to the Fermi paradox and the problem of extraterrestrial life

Millis - Discovery Systems - FDLF 2014

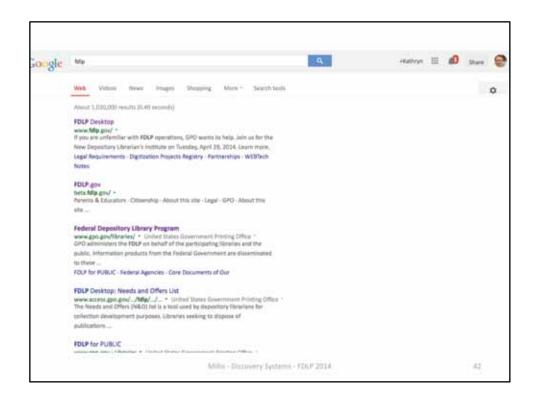
41

At the beginning, I said I'd look at how many gov docs we found, how high they were, and what was higher. This is a good time to look just a little more at what WorldCat considered more relevant that the government document at #42.

"I don't know how to love him" from Jesus Christ Superstar, "All I ask of you" from Phantom of the Opera, "Don't cry for me Argentina" from Evita and "Tell me on a Sunday." Put those together into the Andrew Lloyd Webber anthology and you're #8 in "relevance" for Don't Ask Don't Tell.

In general, I haven't been happy with how WorldCat or Primo rank government documents. I think they end up too low on the result lists, especially when I look at what they ranked as more relevant. But results like this, or the editorials from the Wall Street Journal, which they rank as more "relevant" leave me with even more questions. For instance: what the heck? Do our collections really suck when it comes to government documents (or even just regular books) on these basic, common topics? Are we so lacking in relevant resources? Is searching for 9/11 or biofuels really complex, calling for an advanced search to get good results?

Or do the search tools just not sort very well (yet)?



Not Yet Part 4.

I have compared the results of some basic searches using library search tools, including two discovery systems, to average peoples' use of Google. I think that my students, faculty, staff, & depository users from the general public have Google experiences & expectations in mind when they search Primo or WorldCat, especially if we librarians have given any reason to expect a Google like tool. I think it's likely that they will be disappointed, though I don't know if they'll perceive the problem as our collections, our tools, or their own search skills.

I am not at all pleased at how far down on the result lists gov docs appear, or with some of the stuff that gets called more relevant, or with the absence of CGP. These need not be permanent problems.

I remember when Google was new, with occasionally wrong & bizarre results. When Google Maps was new (and said that my town of 10,000 residents had about fifty each of Burger King and McDonalds restaurants, though we really have just one of each). Google has huge advantages in over a decade of improvement and surveillance. Google knows, when random anonymous people search for 9/11, what they pick. Google knows how long they spend there. Google knows if they go back there. Google knows, because I'm the only person who uses my computer, and because I'm signed into my email, that when I type FDLP I probably mean the Federal

Relevancy ranking

FDIP

- "Temporal resolution analysis in frequency domain linear prediction"
- "Freeze-dried liver powder (FDLP) supports rat hepatocytes in culture"
- "GPO officials put brakes on FDLP overhaul.(newsdesk) (Government Printing Office's Federal Depository Library Program)(Brief article)"
- "Effect of freeze-dried leek powder (FDLP) and nitrite level on processing and quality characteristics of fermented sausages"
- "Direct loans: Federal Direct Consolidation Loan application packet"

Millis - Discovery Systems - FDLP 2014

43

Primo and WorldCat don't have a decade of watching us, as a group or as individuals. In addition to several other problems in ranking it doesn't know whether I want libraries, liver powder, leek powder, or loans.

I imagine that if they get to know me better (and they have world domination as a goal) they might get better at showing me government documents. But I don't know if that will result in other people finding more government documents higher on their results lists.

(I also don't think this is one of the major causes of the problems I've just demonstrated.)

Results, caveats, & questions

- "Relevancy" usually put recent documents lower than date.
 - Historic topics & documents.
- Material more "relevant" than documents questionable.
- Electronic-only, never acquired, & weeded documents not searched so not "discoverable."

Millis - Discovery Systems - FDLP 2014

4.0

My searches were intentionally for recent topics, because I knew I had good ones in the catalog. Sadly, "relevancy" put many lower on result lists than chronology. Had I used historic topics (Agent Orange, the Misery Index, wolves in Yellowstone National Park, World War II), primary sources from the federal government would, if my library had even cataloged them, be far down on a chronological list.

This summer, I'll look at whether discovery tools provide better and easier access to historic documents than the catalog's chronological display does. (But I have to do it before July 1, because after that, I have no catalog.)

That project will also be pretty different because my searches can only find government documents that I have physically in my collection and listed in our catalog. That excludes many older things we never cataloged, but the fact that we did not catalog them is not something I can ask Ex Libris or OCLC to fix.

What I can ask for, and think we and our patrons have a right to, is for electronic documents to be in every discovery tool. As more and more documents get digitized, my patrons ought no longer be limited by cataloging decisions made in the eighties, seventies, and before!

Nor should we need to catalog print copies of things that are freely available on the web.

4. What can we do?

- Widgets
 - http://lpmagnuson.com/test/dlsearch-dugov.html
- · Mt:ngp

Thanks to Lauren Magnuson & Ruth Szpunar!

Millis - Discovery Systems - FDLP 201-

45

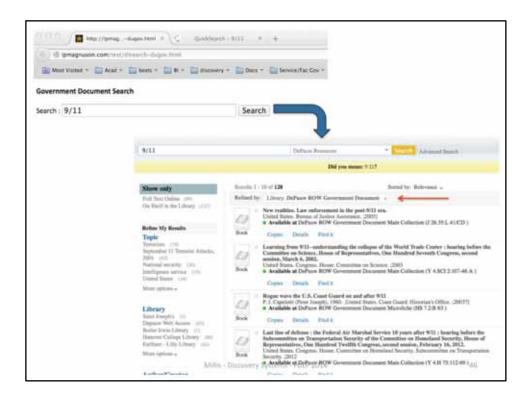
So, finally, Part Four. What can we do?

Here are a couple of quick suggestions for coping with the current situation. First, widgets! Widgets aren't "in the cloud" trendy, but widgets are cool. Widgets, like this one, can provide a search box that's pre-programmed to limit to government documents. They can't help people who search from your home page, but you can add them to LibGuides or web pages.

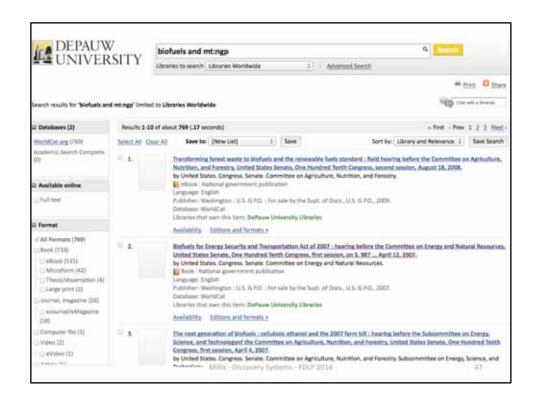
This widget searches my library's Primo, but limits to government documents. (It will die on July 1, when we move to WorldCat.)

Second, with just ten more keystrokes, you can limit your search results to national government publications.

Two colleagues, Lauren Magnuson and Ruth Szpunar helped with this part.



Here's the Primo widget on the upper left, and when you hit "Search," you get the results screen below, which is limited to DePauw's Government Documents collection. It's not perfect, as we have some gov docs in Reference, Science, etc., but it directs patrons to a collection they may not have known about.



To limit a search in WorldCat or other bibliographic databases, type your search word(s) followed by "and mt:ngp"

"And mt:ngp" probably has to be after your search words. That will limit your results to national government publications. You may find things by other nations, but it's a great start.



We can (I'm not saying we should) help library vendors to compile the kind of surveillance data that Google has. What do we use? What do we recommend?

Tools like WorldCat let you make lists, assign star ratings, share via Facebook, Twitter, and almost 300(!) other tools, including Amazon wish lists, Сто Закладок, and Cherry Share.

We can also write reviews for these government documents, enhancing their very brief records with completely honest comments.

Any suggestions for other things that we front line gov doc librarians can do?

What else?

- · Shorter titles!
- Catalog of Government Publications freely available to anybody, including Discovery Layer providers
 - Ex Libris doesn't have
 - · Marcive's Documents Without Shelves
 - Online only.
 - Ebsco, OCLC, & ProQuest do
 - · as subscription
 - · filtering still removes things not held
- Media

Millis - Discovery Systems - FDLF 2014

49

Each of us individually, or all of us linking, starring, and reviewing, can make only the slightest of difference, if any.

It will take larger organizations to make real change.

First, and I realize this is still thinking small, we probably need changes in bibliographic information. For instance, gov docs, especially hearings, need shorter titles. I've seen titles of a hearing by a subcommittee's subcommittee spread across several days that take up half of my laptop's screen! I will be lobbying my congressman.

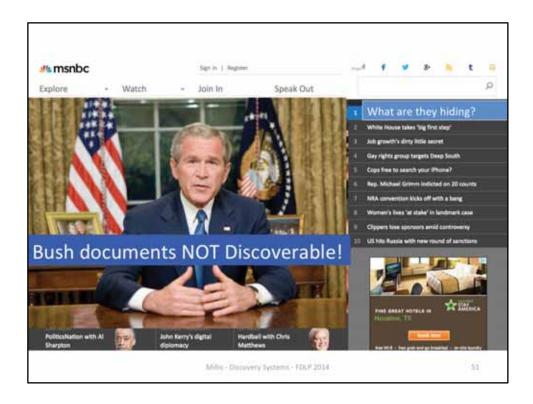
Second, the Catalog of Government Publications needs to be freely available to anybody who wants to do something with it, including the makers of all of the competing discovery layers.

Marcive, Ebsco, OCLC, ProQuest, and other vendors make money because it's not freely available, so I assume that their owners, employees, and lobbied congressmen would disagree. GPO and the FDLP are about access, not profit.

If that isn't easily done, we might need to turn to the media. We need a "study," a few press releases, and a twitter hashtag. My first thought, #NOTdiscoverable.



What is Obama hiding? Why are his documents not in library Discovery tools? I am outraged! #NOTdiscoverable



What is Bush hiding? Why are his documents not in library Discovery tools? I am outraged! #NOTdiscoverable

Google data

 The Value of Google Result Positioning https://chitika.com/google-positioning-value

Millis - Discovery Systems - FDLP 2014

- 53

Once credit, besides Lauren & Ruth, is that Chitika study.

Discussion!?

Kathryn C. Millis millisk@depauw.edu 0172

https://sites.google.com/a/depauw.edu/ fdlp-2014/

Millis - Discovery Systems - FDLP 2014

53

Questions? Comments? Discussion?

Here's my contact info.