

ACRL
Delaware
Valley
Chapter

19 November 2004



Alan Manifold

Systems Implementation Manager

Purdue University Libraries



NOT!



Just a Techie,
ENCompass only

WHO AM I?



🗣️ Google, et al

Pros & cons, new stuff, failure rate

🗣️ Promise/Vision of Federated Searching

Definitions, screen shots

🗣️ Reality of Federated Searching

Problems, successes, challenges

🗣️ Some Reflections

OUTLINE



What does Google provide that libraries have not traditionally provided?



Web

Results 1 - 10 of about 2,350,000 for google not libraries. (0.29 seconds)

Ranking – keeps track of popularity

... the order of records returned is based on the number of links to it on the web ...

Relevancy – analyzes the content

... the order of records is also based on their relevance, determined from the content ...

Really easy to use

... just type in a word, **Google** does the rest ...

Comprehensive

... it hardly matters what topic or area you want, **Google** has scads of hits for it ...

Consistent in response

... **Google** basically always returns the same kind of screen ...

Nice looking

... it's colorful, it's got a cute name, it has indentations and headings, etc ...

GOOGLE – PROS



Google's biggest failings:

Selection: No selection criteria to insure quality, accuracy, peer review, etc.

Authority control: No control over content, which can be inaccurate and inconsistent, as well as being filled with words added just to affect relevancy score

Currency: Returns countless dead links

Weighting: Can weight HTML title tags more than other content, but can't do much beyond that

Cataloging: No descriptive metadata to describe document content overall

GOOGLE – CONS





National Park Service
U.S. Department of the Interior

ParkNet

National Park Service
<http://www.nps.gov/>

You have reached either an unknown or expired web page address. Please check the URL address again, and/or visit our main page at: <http://www.nps.gov/> for navigation assistance. We apologize for any inconvenience this may cause.

10/19/2001

Still listed
more than
two years
after
going 404

Dead links seem like a serious problem,
but nobody seems to care, right?
What can we learn from that?

GOOGLE'S FAILURES



Google is best at fact finding at a shallow depth:
Who starred in that?
When did that happen?
What's its atomic number?
Where can I get one?
Who makes those?
What are its side effects?
How do you do that?
When is it showing?



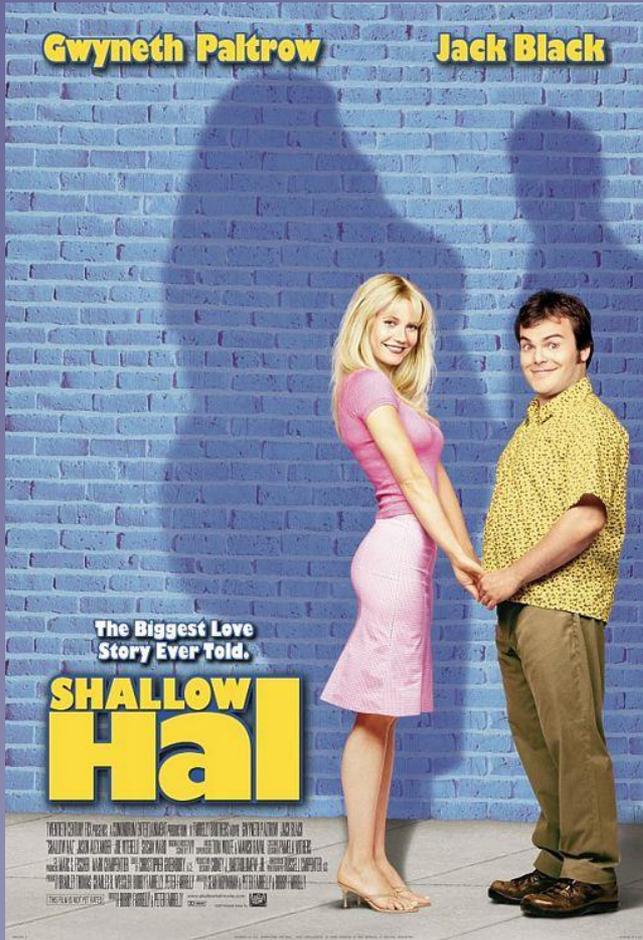
Google is not so good at depth and analysis:
Tell me about it.
Why did they?
How are they related?
Which is better?
What do the experts think?
Has that been proven?
How do we know that?
What led up to that?

SHALLOW vs DEEP



the promise and reality of federated searching

We worry when our children date someone we think is superficial and shallow.



We worry about our patrons when they rely totally on Google for their information needs when it does not provide the depth and quality of information they could be getting.

GOOGLE ADDICTION?



the promise and reality of federated searching

Initially, federated searching was simply the ability to search a number of disparate resources with a single search.

Current federated searching products often also include support for:

- ✓ Multiple protocols
- ✓ Multiple data formats
- ✓ Multiple search types
- ✓ Results consolidation
- ✓ Record de-duping
- ✓ Results sorting



FEDERATED SEARCH



In the future, federated searching products could expand into powerful searching tools.

Already they are evolving towards the ability to support such things as:

-  Searching and presenting video and audio files
-  Personally contoured searching
-  Searching numerical data
-  Searching and presenting non-textual data (e.g. maps, genomes, chemical compounds)
-  Institutional repositories

VISION



These features rely on searching object metadata. As the ability evolves to search object attributes directly, and peripheral options expand for presentation, federated searching systems should be there:

Search for and present aromatherapy solutions

Search webcams by image attributes

Search and reproduce holographs

Search (and even clone) genomes

Search and reproduce chemical compounds

Search parts and assemblies by shape



DISTANT FUTURE?



Federated Searching is a more powerful tool than Google in some significant ways:

Federated Searching

Dynamic

Multiple protocols

Open

Focused



Google

Static

HTTP protocol

Proprietary

Unfocused



IS FS THE ANSWER?



Records in most library systems are coded in MARC, which works well for describing traditional library material content, but is difficult to extend to other material types:

OCLC MARC Bib Record in Raw Form:

00734cam 22002411 45*0001001300000003000600013005001700019008004100036
0100017000770400023000940430012001170500016001290820013001450920019001
5804900090017710000250018624500980021126000570030930000390036635000090
0405504003000414651004800444*ocm00442080 *OCoLC*19940620065418.0*7010
12s1968 pauab b 000 0 eng * ‡a 68021623 * ‡aDLC‡cDLC‡dOCL‡dIPL*‡an-us
---*0 ‡aJK2556‡b.E2* ‡a325.3/73* ‡a325.373‡bEb61f* ‡aIPL1*1 ‡aEblen, Jack Eri
cson.*14‡aThe first and second United States empires;‡bgovernors and territorial gover
nment, 1784-1912.*‡a[Pittsburgh]‡bUniversity of Pittsburgh Press‡c[1968]* ‡aviii, 34
4 p.‡billus., map.‡c24 cm.* ‡a8.95* ‡aBibliography: p. 321-333.* 0‡aUnited States‡x
Territories and possessions.**

MARC



Federated searching products use XML, which codes information for content, not for display.

<MARC>

<MRleader>02412naa 2200289 4500</MRleader><MR001>14582694</MR001>

<MR008>200410e20041001xxu####e###j#####eng#d</MR008>

<MR022><MR022a>0003-3472</MR022a></MR022>

<MR072><MR072a>Article</MR072a></MR072>

<MR100 ind1="1" ind2="0"><MR100a>O'Donnell, Ryan P.</MR100a></MR100>

<MR700 ind1="1" ind2="0"><MR700a>Ford, Neil B.</MR700a></MR700>

<MR700 ind1="1" ind2="0"><MR700a>Shine, Richard</MR700a></MR700>

<MR245 ind1="1" ind2="0"><MR245a>Male red-sided *snip*</MR245a></MR245>

<MR270><MR270a>Dept of Zoology, Oregon State U</MR270a></MR270>

<MR514><MR514a>Peer Reviewed</MR514a></MR514>

<MR520><MR520a>...snip...</MR520a></MR520>

<MR654><MR654a>GARTER snakes</MR654a></MR654>

<MR773>

<MR773t>Animal Behaviour</MR773t><MR773g> Oct2004, Vol. 68 Issue 4, p677</MR773g>

</MR773>

<MR903><MR903a>20041001</MR903a></MR903>

<MR945>

<MR945m>68</MR945m><MR945n>4</MR945n><MR945p>677</MR945p>

</MR945>

</MARC>

XML



MegaSearch Databases by Title

Quick Search MegaSearch Databases by Subject MegaSearch Databases by Title All Databases by Subject All Databases by Title Help Exit

Keyword Anywhere [] All of These []
AND Keyword Anywhere [] All of These []
AND Keyword Anywhere [] All of These []
10 Records per Page
Return Results in [Default] order
[Search]
 Remove Duplicate Hits?

EnCompass
federated
search by title
screen

Select search options

Check databases below to search them simultaneously
B - C - D - E - F - G - H - I - J - K - L - M - N - O - P - Q - R - S - T - U - V - W

MegaSearch Databases

Include Purdue Libraries catalog in search?

Select	Database or Subject	Summary - Display Full Descriptions	Coverage	Full Text	More
<input type="checkbox"/>	ABI/Inform Global Access	Use to find journal articles in business and management.	Mid-1970's - present	Partial	
<input type="checkbox"/>	Academic Search Elite	General information from journals, magazines, and newspapers in all subject areas.	1984 - present; fulltext, 1990 - present.	Partial	
<input type="checkbox"/>	Aerospace Database	Aeronautics, Astronautics, Space Sciences, and supporting fields.	1986 - present	Partial	
<input type="checkbox"/>	AGRICOLA	Agriculture and allied disciplines	1979 - present	No	
<input type="checkbox"/>	America: History and Life	1964 - present	Partial		

Select databases to search

SAMPLE SCREEN #1



MegaSearch Databases by Subject - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Favorites Refresh

Address http://vader.lib.purdue.edu:20018/logicrouter/servlet/LogicRouter?OUTPUTXSL=msdbssubj.xml&hier=h1.12.1.5&tree=c&api_1=COLL_GET_CHILDREN&#h1.12.1.5 Go Links >>

Saved Searches
Please Login

Saved Records
Please Login

Subjects and Databases

Include Purdue Libraries catalog in search?

Expand

+ Arts and Humanities

Expand

+ Business and Economics

Expand

- Engineering and Technology [Select All](#) [Deselect All](#)

Select	Database	Summary - Display Full Descriptions	Coverage	Full Text	More
<input type="checkbox"/>	Aerospace Database	Aeronautics, Astronautics, Space Sciences, and supporting fields.	1986 - present	Partial	i
<input type="checkbox"/>	Applied Science and Technology Abstracts	Industrial and mechanical arts topics. *Partial full-text coverage in OmniFile Full Text Mega.	1983 - present	Partial	i
<input type="checkbox"/>	Wiley Encyclopedia of Electrical and Electronics Engineering	Electrical and electronics engineering.		Partial	i

Expand

+ General and News

Expand

Government and Law

Life Sciences

Medicine and Health

Expand

+ Physical Sciences

Expand

+ Social Sciences and Education

Done Internet

Expand subjects

Select DBs

EnCompass federated search by subject screen

SAMPLE SCREEN #2





- Quick Links**
- ▶ Your Library Account
 - ▶ Interlibrary Loan
 - ▶ Course Reserves
 - ▶ Hours
 - ▶ Maps

- About
- Help
- Catalogs
- Libraries and Units
- Articles and Databases
- Guides and Instruction
- Giving
- Ask a Librarian

Searching...

- Quick Search
- MegaSearch Databases by Subject
- MegaSearch Databases by Title
- All Databases by Subject
- All Databases by Title
- Help
- Exit

Retrieving Results

Collection
Academic Search Elite
Lexis-Nexis Academic
NewsBank NewsFile
Newspaper Source
OmniFile Full Text Mega
Proquest Research Library



Search Status (limited to 100 hits per database)
7 Hits (7 Total)
34 Hits (941 Total)
Search Failed
24 Hits (24 Total)
40 Hits (492 Total)
Database Failure

Federated search results in progress

SAMPLE SCREEN #3



1	Search Status (limited to 100 hits per database)	
	7 Hits (7 Total)	
2	34 Hits (941 Total)	
3	Search Failed	5
	24 Hits (24 Total)	
	58 Hits (492 Total)	
4	Database Failure	

- ① Results limited for better response time
- ② We don't know which 34 hits and we can't get the ones after 100 (the limit)
- ③ It got to the database, but the search itself failed
- ④ It didn't connect to the database
- ⑤ Click to get usually useless messages to explain failures (e.g. "Unknown error")

NEEDS EXPLAINING

[Ohio GOP Challenges 35,000 Voters Washington Post, The; 01/23/2004](#)

Source: Newspaper Source

[Search Database Directly](#)

[Analysis: Democratic presidential field on health care MICHELE NORRIS; All Things Considered \(NPR\); 07/07/2003](#)

Source: Newspaper Source

[Search Database Directly](#)

[Democratic hopefuls clash over war, health care Steven Thomma; Knight Ridder Tribune Washington Bureau \(DC\); 05/03/2003](#)

Source: Newspaper Source

[Search Database Directly](#)

[Democrats on Stump Plot Their War Rhetoric Glen Johnson; Boston Globe, The \(MA\); 03/11/2003](#)

Source: Newspaper Source

[Search Database Directly](#)

[Igniting the Black Base. Campo-Flores, Arian; Newsweek; 11/1/2004, Vol. 144 Issue 18, p27](#)

Source: Academic Search Elite

[Search Database Directly](#)

[An analysis of the perceptions and characteristics of... Dinnebeil, Laurie A.; Topics in Early Childhood Special Education; Summer98, Vol. 18 Issue 2, p118](#)

Source: Academic Search Elite

[Search Database Directly](#)

[The Atlanta Journal-Constitution](#)

Source:  LexisNexis

[Download](#)

[Search Database Directly](#)

[The Australian](#)

Source:  LexisNexis

[Download](#)

[Search Database Directly](#)

[The Business Times Singapore](#)

Source:  LexisNexis

[Download](#)

[Search Database Directly](#)

[The Daily Telegraph \(Sydney, Australia\)](#)

Source:  LexisNexis

[Download](#)

[Search Database Directly](#)

“Show All” results screen – no sorting specified

SAMPLE SCREEN #4



Database Source	 LexisNexis™
Title	The Australian
Description	November 3, 2004 Wednesday All-round Country Edition, LOCAL; Pg. 1, 848 words, Bush and Kerry walk a knife-edge, Roy Eccleston, Washington correspondent

[Search Database Directly](#)

[Download](#)

The record display screen shows selected fields from record, as determined by the library. The presence of the  button means there is a URL in the record.

SAMPLE SCREEN #5



Although federated searching has a number of successes to its credit, there are also a number of problems. And since the products are relatively immature, there are many more challenges ahead. Let's look at:

- ☹ Problems
- ☹ Challenges
- ☺ Successes

REALITY



A number of problems contribute to the difficulty of making federated searching match its vision. Among them:

- ☠ Lack of standards
- ☠ Multiple protocol support
- ☠ Multiple data formats
- ☠ Range of vendor support
- ☠ Search definitions
- ☠ Z39.50 problems
- ☠ HTTP Search Engine (HSE) connectors

PROBLEMS



What does a title search cover...

On my online catalog?

On another online catalog?

On an A&I database from vendor P?

On an A&I database from vendor E?

On Amazon or Barnes & Noble?

On Associations Unlimited?

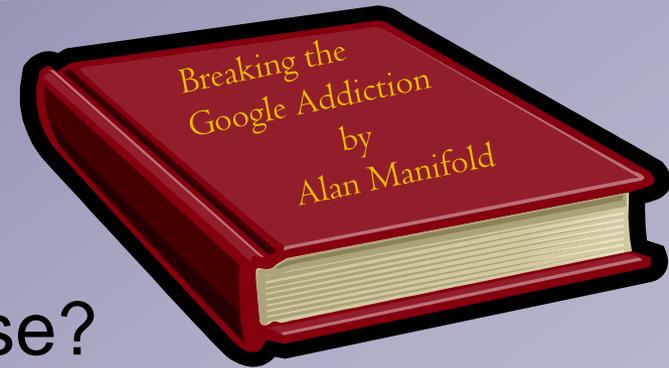
On Google?

On Funding Opportunities Database?

On Reference Suite@FACTS.com?

On Biography Resource Center?

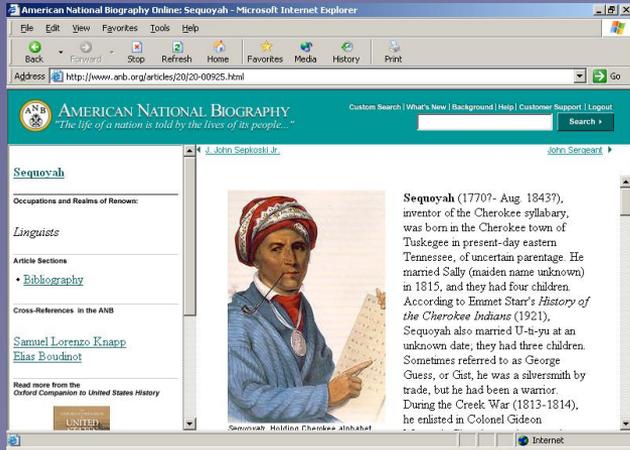
With these Z39.50 Attributes: Use:4, Relation:3,
Position:3, Structure:1, Completeness:1,
Truncation:1?



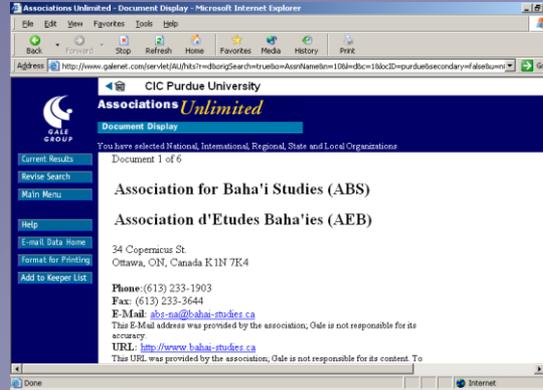
SEARCH & Z PROBLEMS



HSE Connectors have as their mission to extract specific data fields whether they are there or not.



American National Biography

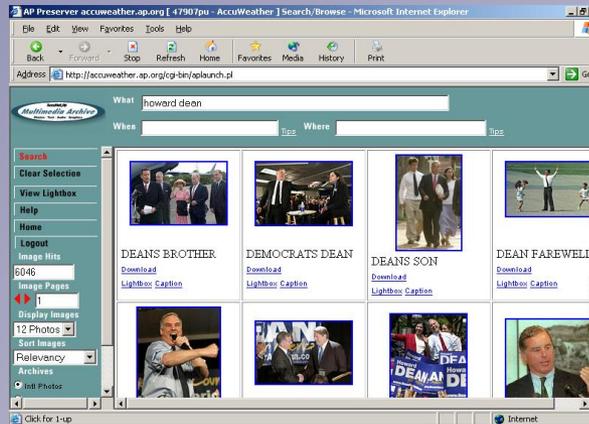


Associations Unlimited

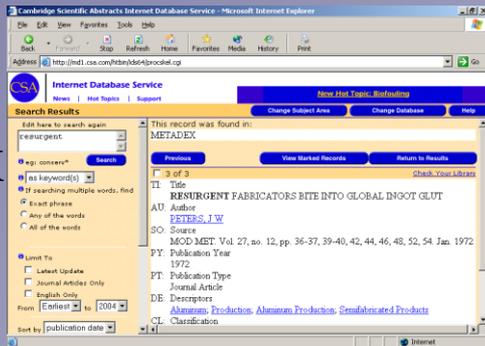
Funding Opportunities



Accunet/AP Photo Archive



AltaVista



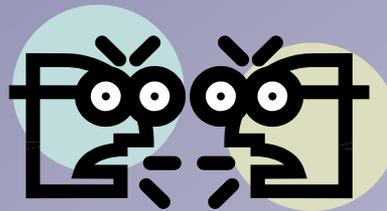
METADEX

HSE CONNECTORS



V Said:

Database vendors complain about HSE connectors because they pound the web servers much more than individual users could do.



Z Said:

System vendors hate using HSE connectors, but the database vendors have not provided alternatives, such as XML gateways or Z-connections.

VENDOR vs VENDOR



Authorization

Make sure only authorized users can get to specific resources

Connectors

Keep current connectors working and move away from HSE connectors to something more stable

Response time

Search multiple resources faster

Integrating new resources

As new protocols and resources come into being, federated search systems need to keep up

De-duping and managing results

When results are like apples and oranges, sorting and de-duping are tough, but the users expect it

CHALLENGES



IT REALLY WORKS!



At Purdue, we have 119 databases listed on our “MegaSearch” pages, about half HSE and half Z39.50

Endeavor claims 138 ENCompass sites

Ex Libris claims 531 MetaLib sites

MuseGlobal MuseSearch (couldn't tell)

Sirsi SingleSearch (couldn't tell)

WebFeat claims 1500 sites

SUCCESSSES



Federated
Searching

Web [Images](#) [Groups](#) [News](#) [Froogle](#) [more »](#)

libraries not google

Search

[Advanced Search](#)
[Preferences](#)

Web

Results 1 - 10 of about 2,350,000 for [libraries not google](#). (0.29 seconds)

[Ranking](#)

... need to become sophisticated enough to figure quality of results ...

[Relevancy](#)

... need to analyze the content enough to determine relevance ...

[Really easy to use](#)

... single box searching with advanced options available, maybe? ...

[Comprehensive](#)

... I think we want to distinguish what fs is for, rather than trying to cover it all ...

[Consistent in response](#)

... put in a search, get records back, same procedure no matter what the result ...

[Nice looking](#)

... concentrate on user interface design principles to get something attractive ...

To attract our patrons to use federated searching, we need to address Google's strengths head-on.

IS THIS ENOUGH?



Maybe we should also look at the problem differently:



We try to get people to come to the library, but maybe a better model for the web would be to put what the library offers into places people are already going. What would this look like?

PERSPECTIVE



Both Google and Yahoo now include links to WorldCat. Type “find in a library” with any search to get library info:

Find in a Library powered by WorldCat

Libraries in your region

House of Bush, house of Saud : the secret relations powerful dynasties

By: Craig Unger
 Type: English : Book : Non-fiction
 Publisher: New York : Scribner, ©2004.
 ISBN: 074325337X

Find this item in a library:

Enter a postal code, state, province or country

Regional Libraries

Web Images Groups News Froogle [more »](#)

[Advanced Search](#)
[Preferences](#)

Web

Tip: Try removing quotes from your search to get more results.

[Find in a Library: House of Bush, house of Saud : the secret ...](#)
House of Bush, house of Saud : the secret relationship between the world's two most powerful dynasties. By: Craig Unger. Type: English : Book : Non-fiction. ...
www.worldcatlibraries.org/wcpa/ow/e359b1bb355271aaa19afeb4da09e526.html - [Similar pages](#)

[Find in a Library: American dynasty : aristocracy, fortune, and ...](#)
American dynasty : aristocracy, fortune, and the politics of deceit in the house of Bush • By: Kevin P Phillips • Publisher: New York : Viking, 2004. ...
www.worldcatlibraries.org/wcpa/ow/9c5865dea9047e87a19afeb4da09e526.html - [Similar pages](#)

Library Name	City	State/Province Country	Library Info	Library Type
Alexandrian Public Library	Mount Vernon	Indiana		Public
Allen County Public Library	Fort Wayne	Indiana		Public
Anderson Public Library	Anderson	Indiana		Public
Ball State University	Muncie	Indiana		Academic
Carmel Clay Public Library	Carmel	Indiana		Public

OCLC AND GOOGLE





The bottom line is value.
We need to give our patrons
valuable services in exchange
for their time and effort.

Federated Searching
adds value to the library's
offerings. Patrons get
more results per minute
spent.



VALUE

Breaking the
Google
Addiction
*the promise and reality of
federated searching*

Picture a patron doing federated searching and getting one-click access to:

- 🖱 Full text of journal articles
- 🖱 Photos, graphs, maps, video clips, sounds
- 🖱 E-Books full text
- 🖱 Data sets in an institutional repository
- 🖱 Selected peer-reviewed websites
- 🖱 Dictionaries and encyclopedias
- 🖱 Specialized software

ENVISION



When we deliver that, our patrons will rush to choose federated searching over Google.



WE SHALL PREVAIL!

