



Metasearching Lessons from the California Digital Library



Roy Tennant

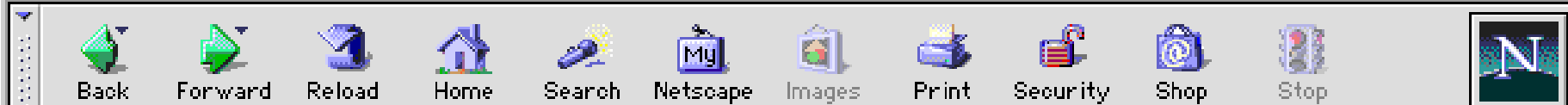
California Digital Library

Metasearching at CDL: A Brief History

- 1997, UC San Diego creates *Database Advisor*
- 1999, CDL decides to create a metasearch tool based on DBA, to be called SearchLight
- 2000, SearchLight is launched



CALIFORNIA DIGITAL LIBRARY

Netsite: <http://searchlight.cdlib.org/cgi-bin/searchlight?Science>

What's Related



CALIFORNIA DIGITAL LIBRARY

COLLECTIONS & SERVICES

GUIDES

NEWS

ABOUT THE CDL

[home](#) • [search](#) • [feedback](#)

Welcome to SearchLight - Sciences/Engineering

Go to: [CDL](#)

With SearchLight, you can search many of the databases and other resources available to CDL users — all at the same time. It can find books, journal and encyclopedia articles, and quality Internet sites. SearchLight will run the search and bring back the results. With SearchLight, you don't have to worry about finding the best databases first — it does it for you. For social sciences and humanities topics, go to [SearchLight - Social Sciences/Humanities](#).



SearchLight will search ALL science/engineering databases in the CDL profile. If you wish, you can target your search to a more specific list of databases by selecting the specific subtopics at the right: [\[Explain\]](#)

Enter your search terms:

search >

*for example: hiv protease inhibitor*How many minutes are you willing to wait? [\[Explain\]](#)☒ **ALL Science/Engineering** [\[more info\]](#)☐ **Biological & Medical Sciences**☐ **Engineering & Technology**☐ **Physical Sciences**

Search Light

SearchLight
search in many

would be found in each one, ranked from the most to least hits. You can then link directly to the database of your choice to perform your search. [More help](#)

<http://searchlight.cdlib.org/cgi-bin/searchlight>



CALIFORNIA DIGITAL LIBRARY



COLLECTIONS & SERVICES

GUIDES

NEWS

ABOUT THE CDL

[home](#) • [search](#) • [feedback](#)

SearchLight

Sciences/Engineering

You searched for: **tamoxifen**[Refine your search?](#) | [Help](#) | [Feedback](#)

UC San Francisco

Go to: [UCSF Library](#)

Remember — not everything is available online. There are also a number of print resources in the UCSF Library that can lead you to useful information. See [important print resources at UCSF](#).



To view actual search results in new browser windows, select either the links under **# Hits** or **Resource Name**.

Jump to:

[Books](#)

Library catalogs and other book citations

[Journal Indexes](#)

Indexes and abstracts to journal articles

[Electronic Journals](#)

Full-text online journals

[Full-Text Resources](#)

Electronic texts and documents

[Reference Resources](#)

Encyclopedias, dictionaries, directories

[Web Directories](#)

Links to Internet resources

Books

- Library catalogs and other book citations

Jump to: [Books](#) | [Journal Indexes](#) | [Electronic Journals](#) | [Full-Text Resources](#) | [Reference Resources](#) | [Web Directories](#)

Hits

Resource Name & Description

35

[WorldCat](#) [\[more info\]](#)

Citations for books, journals and materials in other formats

16

[MELVYL Catalog](#) [\[more info\]](#)

Library catalog for all UC campuses

Journal Indexes

- Indexes and abstracts to journal articles

Jump to: [Books](#) | [Journal Indexes](#) | [Electronic Journals](#) | [Full-Text Resources](#) | [Reference Resources](#) | [Web Directories](#)

Hits

Resource Name & Description

Journal IndexesJump to: [Books](#) | [Journal Indexes](#) | [Electronic Journals](#) | [Full-Text Resources](#) | [Reference](#)**- Indexes and abstracts to journal articles**[Resources](#) | [Web Directories](#)

# Hits	Resource Name & Description
--------	-----------------------------

8767	BIOSIS Previews [more info]
----------------------	---

Life sciences and biology; indexes over 6000 journals, plus books and conferences

6787	MEDLINE/HealthSTAR [more info]
----------------------	--

Citations to journal articles in all areas of medicine and health sciences

2229	Current Contents [more info]
----------------------	--

Covers 6500 scholarly journals in all disciplines from 1989+

278	MAGS: Magazine & Journal Articles [more info]
---------------------	---

Citations to general-interest articles, many with abstracts, some with full text

Electronic JournalsJump to: [Books](#) | [Journal Indexes](#) | [Electronic Journals](#) | [Full-Text Resources](#) | [Reference Resources](#) | [Web](#)**- Full-text online journals**[Directories](#)

# Hits	Resource Name & Description
--------	-----------------------------

678	Elsevier ScienceDirect [more info]
---------------------	--

Electronic journal articles published by Elsevier

450	Academic Press (IDEALIBRARY) [more info]
---------------------	--

Electronic journal articles published by Academic Press

324	Kluwer Online [more info]
---------------------	---

Electronic journal articles published by Kluwer Academic Publishers

45	JSTOR [more info]
--------------------	---

Electronic journal articles from many publishers provided by JSTOR

Full-Text ResourcesJump to: [Books](#) | [Journal Indexes](#) | [Electronic Journals](#) | [Full-Text Resources](#) | [Reference Resources](#) | [Web](#)**- Electronic texts and documents**[Directories](#)

Metasearching at CDL: A Brief History

- 2001-3, CDL looks beyond the limitations of the SearchLight service
- 2002-3, Commercial products become viable
- 2003, CDL Metasearch Infrastructure Project formally launched



CALIFORNIA DIGITAL LIBRARY

Lessons from SearchLight

- *Metasearching is worth doing* (it's what many users want)
- For a large research library, *metasearching is best focused on particular needs or subject areas* (the “drinking from a firehose” problem)
- *Not all databases are created equal* (e.g., we need a way to focus on core databases, or have results from core databases ranked higher than others)
- Focusing on specific audiences or needs *provides an opportunity to expand service beyond simple searching*



Simple Search

[Help](#)
[Ask a Librarian](#)
[Feedback](#)
[KU Libraries](#)
[Reset](#)
[Exit](#)

Quick Navigation: Go to...

Search for:
 20 Records per Page [search](#)
[Simple Search](#) | [Advanced Search](#)

Quick Start

To search:

1. Check one or more collections at the right
2. Enter words to search in the box above
3. Click the "search" button

Login

COMING -- the ability to log in and then personalize your session by (1) creating a custom group of databases; (2) saving searches; and (3) saving records.

This service was funded in part by KU students through the Tuition Enhancement Program

☐ Include KU Libraries' main catalog in search?

Expand	Select	Resource
	<input type="checkbox"/>	Area and Cultural Studies View Description
	<input type="checkbox"/>	Arts and Humanities View Description
	<input type="checkbox"/>	Business View Description
	<input type="checkbox"/>	Current Events and News View Description
	<input type="checkbox"/>	General and Reference View Description
	<input type="checkbox"/>	History View Description
	<input type="checkbox"/>	Internet Search Engines View Description
	<input type="checkbox"/>	Languages and Literature View Description
	<input type="checkbox"/>	Library Catalogs - KU View Description
	<input type="checkbox"/>	Medicine and Health View Description
	<input type="checkbox"/>	Science and Engineering View Description
	<input type="checkbox"/>	Social Sciences View Description
	<input type="checkbox"/>	all Digital Library Databases A-Z View Description

All KU Libraries Electronic Resources: A B C D E F G H I J K L M N O P R S T U V W Z

Please note: Not all KU Libraries electronic resources are available yet in the Digital Library

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**

Expand

**General and News**[Select All](#)[Deselect All](#)

Select	Database	Summary - Display Full Descriptions	Coverage	Full Text	More
<input type="checkbox"/>	Academic Search Elite	General information from journals, magazines, and newspapers in all subject areas.	1984 - present; fulltext, 1990 - present.	Partial	i
<input type="checkbox"/>	AccuNet/AP Photo Archive	News photos from the 1840's to today are covered. this resource is updated continuously.	1840's - present	Partial	i
<input type="checkbox"/>	American National Biography Online	Profiles of 18,000+ men and women who have influenced American history and culture.	colonial-present	Partial	i
<input type="checkbox"/>	ArticleFirst	Business, Humanities, Medicine, Popular Culture, Science, Social Science, Technology.	1990 - present	Partial	i
<input type="checkbox"/>	Biography Resource Center	Worldwide coverage of past and present noteworthy people. Narrative biographies, key facts, selected articles, and more.		Partial	i
<input type="checkbox"/>	Ethnic NewsWatch	The ethnic, minority, and native press in America.	1985 - present	Partial	i
<input type="checkbox"/>	Funk and Wagnall's New World Encyclopedia	General Encyclopedia.		Partial	i
<input type="checkbox"/>	Informe (English)	Hispanic reference material covering cultural and lifestyle issues, health and parenting, entertainment, politics and current events.	1996 - present	Partial	i
<input type="checkbox"/>	JSTOR - The Scholarly Journal Archive	Electronic access to older scholarly journals.	Varies	Partial	i

working papers
journals
books
policy briefs
case studies
course packs
schedule of events
maps/country data
site index

about
reviews
remote access
subscribe
access help

© columbia
university press

CIAO Focus, January 2004: India and Pakistan

This month's meeting between the leaders of India and Pakistan has renewed hopes that a real and lasting peace can be made between the two nuclear powers. Prime Minister Atal Behari Vajpayee and Pakistani President Pervez Musharraf put forth a new initiative aimed at normalization of relations. In a sentiment echoed by Musharraf, Vajpayee called for a bold transition. "We have to change South Asia's image and its standing in the world," Vajpayee said. The dispute over Kashmir, which has sparked two of their three previous wars, remains a major stumbling block. Separatist violence continues to be on the rise in Kashmir, with over a dozen people killed in December despite a truce. The recent rapprochement has been greeted with optimism by the world community.

From CIAO's database:

[Finding a Kashmir Settlement: The Burden of Leadership](#)

[The Start of a Beautiful Friendship? The United States and India](#)

[Is South Asian Nuclear Politics Different?](#)

[Some Cooperative Initiatives for India-Pakistan Dialogue and Risk Reduction](#)

[Why India Went Nuclear](#)

[Politics in Pakistan Post-September 11, 2001](#)

Outside Links:

[United Nations Military Observer Group in India and Pakistan](#)



Map of Kashmir Region

search ciao

☐ Working Papers
☐ Journals ☐ Policy Briefs
☐ Books ☐ Case Studies

All subjects ▼

All Dates ▼

search

[help](#) | [advanced](#)

featured content

[The National Academy of Public Administration](#)

[Asia Society](#)

[CIAO Atlas Pages on India](#)

[CIAO Atlas Pages on Pakistan](#)

Principles

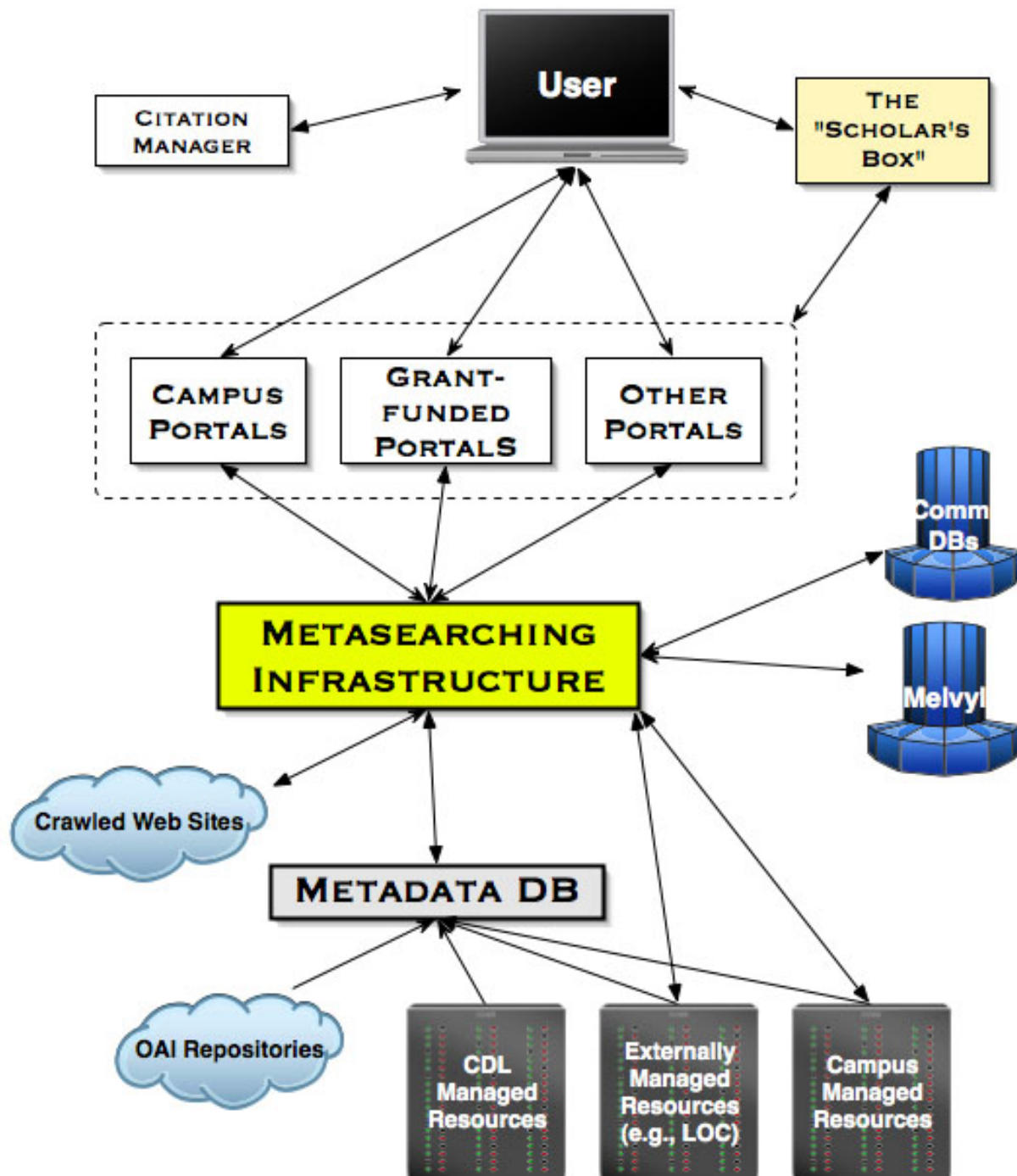
- Metasearching is a tool that is appropriate for some tasks, but not for others
- Only librarians like to search, everyone else likes to *find*
- All things being equal, one place to search is better than two or more
- “Good enough” is often just that

Principles

- The size of the result set isn't as important as how the results are displayed (the Google lesson)
- Our ability to create effective one-stop searching is dependent on our ability to appropriately target user needs
- Services should be placed as close to the user as possible

Building a Metasearch Infrastructure

- Goal: to build a reference implementation that allows us to “eat our own dog food” while providing:
 - An implementation for campuses that don’t want to mess with it
 - An easy-to-modify set of templates for those who do
- Requirements: shared resource descriptions, templates, implementation guidelines, etc.



Collection Development
Process

Collections

Projects, Programs &
Services

Committees & Groups

Digital Library Building
Blocks

Instructional Materials

Assessment

News & Events

Report a Problem

Contact the CDL

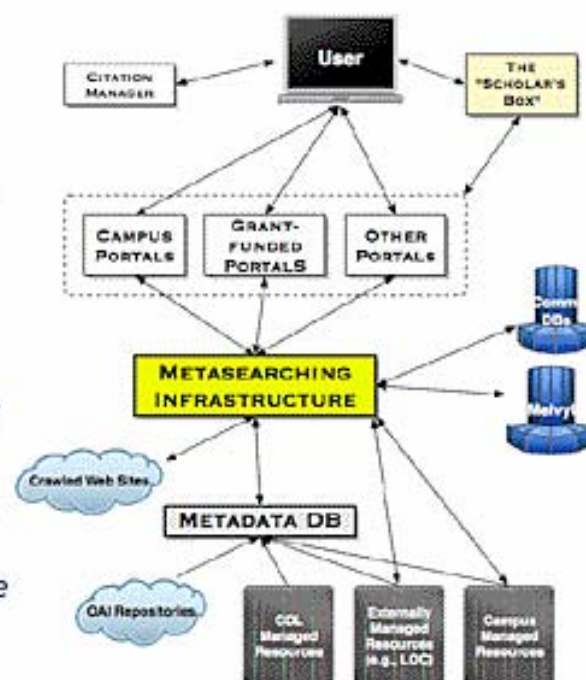
[Projects, Programs & Services](#)

The Metasearch Infrastructure Project

The Metasearch Infrastructure Project aims to create an infrastructure that can be used by campus libraries and CDL to tailor search portals to particular user needs. Metasearching has also been called "cross-database searching" and "federated searching". This infrastructure would replace the existing **SearchLight** service with a "family" of portals aimed at particular audiences and needs. Based on a thorough review of the marketplace, we are presently working to acquire a commercial product by first quarter 2004. Contact **Roy Tennant** for more information.

Our vision of how this tool could be deployed within the university information environment is depicted in the diagram to the right (click on it for a larger version). The "**Scholar's Box**" refers to a tool being

METASEARCHING INFRASTRUCTURE SCHEMATIC



Roy Tennant, 11 Dec 2003

www.cdlib.org/inside/projects/metasearch/

ch/

Done

our metasearching infrastructure as a resource for university faculty, K-12 teachers, and others to use to create learning objects, exhibitions, and other types of products based on selected and

EZArticleFinder

Search less, find more...

[More search options](#) | [Search tips](#)

Search for **giant squid** found 60 results. 1-50 of 60 records retrieved.

Best bets for finding articles related to **giant squid**

- * [BIOSIS Previews](#)
- * [Expanded Academic ASAP](#)
- * [Lexis-Nexis](#)

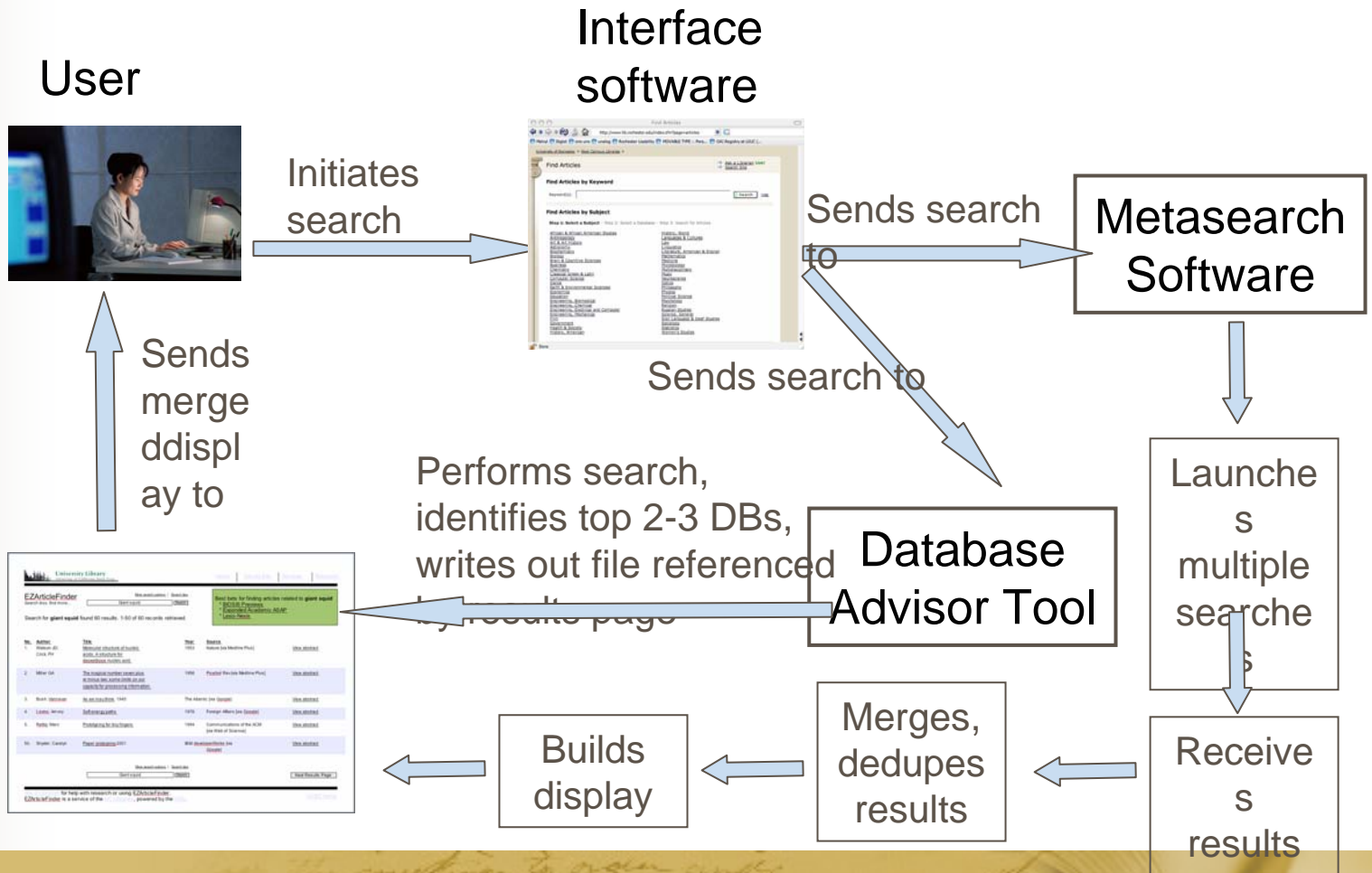
No.	Author	Title	Year	Source	
1.	Watson JD; Crick, FH	<u>Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid.</u>	1953	Nature [via Medline Plus]	View abstract
2.	Miller GA	<u>The magical number seven plus or minus two: some limits on our capacity for processing information.</u>	1956	Psychol Rev [via Medline Plus]	View abstract
3.	Bush, Vannevar	<u>As we may think</u>	1945	The Atlantic [via Google]	View abstract
4.	Lovins, Amory	<u>Soft energy paths</u>	1976	Foreign Affairs [via Google]	View abstract
5.	Rettig, Marc	<u>Prototyping for tiny fingers</u>	1994	Communications of the ACM [via Web of Science]	View abstract
50.	Snyder, Carolyn	<u>Paper protoyping</u>	2001	IBM developerWorks [via Google]	View abstract

[Next Results Page](#)

[Ask a Librarian](#) for help with research or using EZArticleFinder.
EZArticleFinder is a service of the [UC Libraries](#), powered by the [CDL](#).

[UCSC home](#)

Database Advisor Service



Plenty of Problems to Go Around

- Database Vendor Issues
- Software Provider Issues
- Library Issues
- User Issues

Database Provider Issues (the short list)

- Access control (robust authentication and authorization)
- Load
- Inappropriate searches (searching databases that don't apply)
- Branding and “unfair” deduping

Software Provider Issues (the short list)

- Access management
- Search mapping
- Unreliability of targets
- Systems that don't support an API (that must be screen-scraped)
- Inadequate result data for good:
 - Deduping
 - Ranking

Library Issues (the short list)

- Selecting the right system
- Cost (both upfront and ongoing)
- System design and implementation
- System maintenance
 - Ability to add new resources/targets
 - Ease of interface changes and upgrades
- System inadequacies (e.g., ranking)

User Issues (the short list)

- What must I go through before hitting the search button?
- How difficult is it to review results?
- Are results ranked by relevance? (that will be my assumption)
- Will I get buried? (too many sources, too many results?)
- Do I have methods to easily focus in on what I want?
- Once I find what I want, can I get to the full-text with a click?
- Can I copy a citation and put it in my paper?

Challenges

- Software still needs improvement (understatement)
- Some databases are still not searchable
- How do we create an infrastructure that is easy to deploy for a variety of purposes and/or audiences?
- We can learn from other systems (e.g., redlightgreen)?
- Standards are on the way (e.g., NISO), but many of us have deployed systems that could use solutions *now*

Metasearching Today & Tomorrow

■ Today:

- One-stop shopping
- Broad subject categories with many databases each
- Not integrated with any other system (e.g., course management systems)

■ Tomorrow:

- Tailored portals for specific user needs or topic areas
- Targets created by libraries for specific purposes (e.g., focused crawling of web sites, harvesting of repositories)
- Dynamic selection of sources based on user