

# Improving Search for Government Resources: Metasearch at the Library of Congress

*Chris Testa, Web Services Customer Requirements Manager*  
*David Woodward, Systems Analyst*

# Metasearch

# The Problem



## LIBRARY CATALOGS

Print, multimedia and online resources



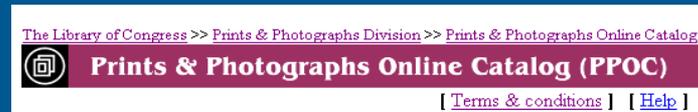
## AMERICAN MEMORY

Historic maps, photos, documents, audio and video



## THOMAS

Legislative information: current and historical



# Metasearch

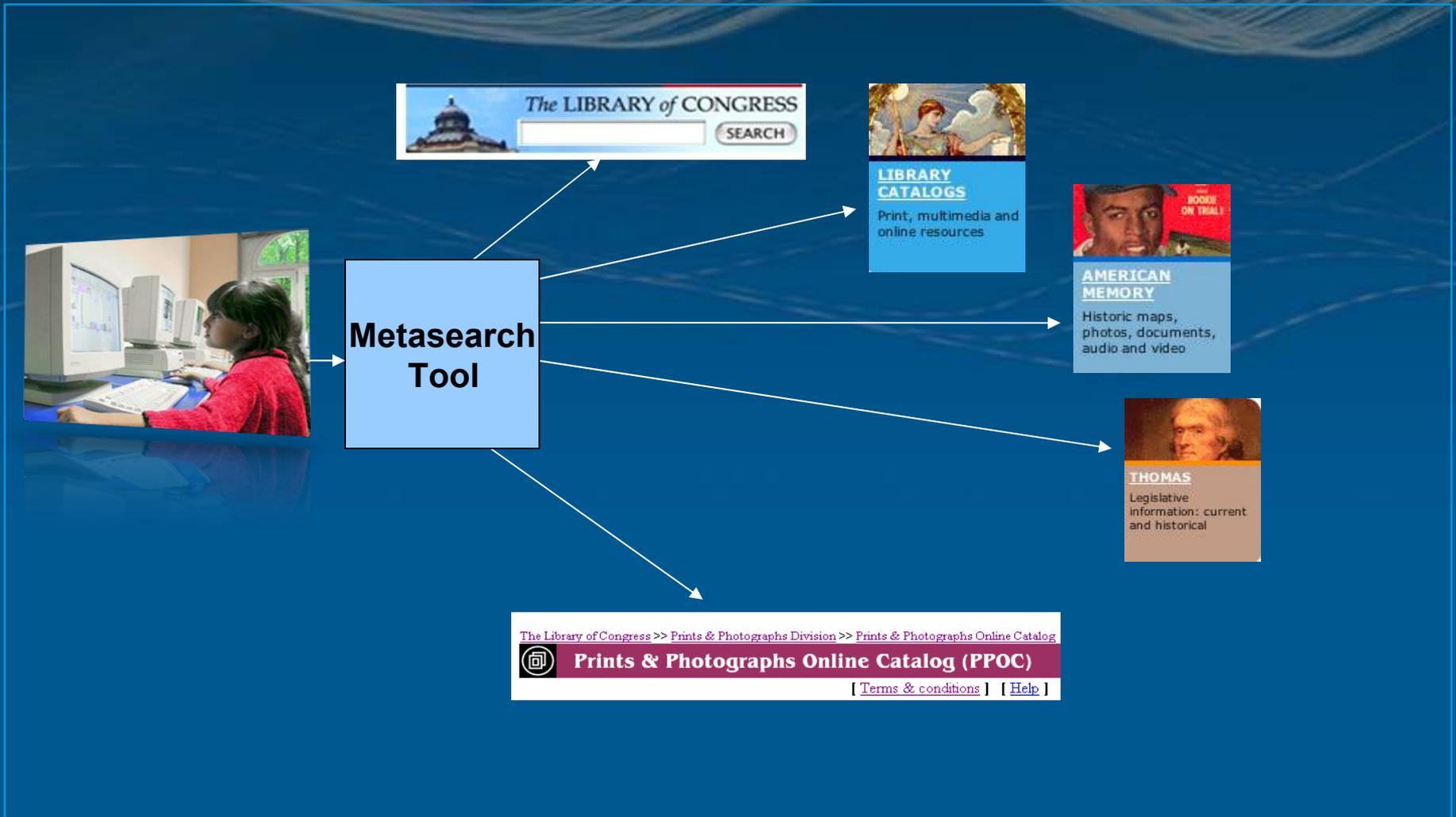
# Why Metasearch?

- Improve discovery: current search box web pages only
- Short development time, inexpensive tool
- Learning experience, exposing data/search issues



# Metasearch

# The Model



# Metasearch

# Homepage

The LIBRARY of CONGRESS

ASK A LIBRARIAN | DIGITAL COLLECTIONS | LIBRARY CATALOGS

The Library of Congress > New Search Home



[New Search Home](#)  
[FAQs](#)  
[Search Help](#)

## New Search (BETA)

For the first time you can search the largest sections of the Library's site from one search box. [Please let us know what you think.](#) Thank you for helping us to improve this service.

*Examples: Thomas Jefferson, copyright, "John Coltrane" "solar energy" (Use quotes for exact matches.)*

### Select Sources to Search

[Select all](#) | [Clear All](#)

- U.S. historical and cultural collections (American Memory)
- Library of Congress Online Catalog
- Prints & Photographs Online Catalog
- Legislation, Congressional Activity (THOMAS Legislative Information System)
- Library of Congress Web site

[SEARCH](#)

[About](#) | [Site Map](#) | [Contact](#) | [Accessibility](#) | [Legal](#)

# Metasearch

# Initial Search Results

The LIBRARY of CONGRESS [ASK A LIBRARIAN](#) [DIGITAL COLLECTIONS](#) [LIBRARY CATALOGS](#)

[The Library of Congress](#) > [New Search Home](#) > Search Results

**Search Again**

**Select Sources:**

- U.S. historical, cultural collections
- Library of Congress Online Catalog
- Prints & Photograph Online Catalog
- Legislation, Congressional Activity
- Library of Congress Web site

[Search](#)

[New Search Home](#)

[FAQs](#)

[Search Help](#)

---

## Search Results

### U.S. historical, cultural collections (at least 500 results)

1. [Silent Film Clip of Captain Pearl R. Nye](#)  
moving image
2. [\[Thomas Jefferson, President of the United States\]](#)  
Tiebout, Cornelius, 1777-1830, engraver. | [1801(?)] | still image
3. [Thomas Jefferson, third president of the United States](#)  
Pendleton's Lithography. | [1828(?)] | still image

[See more results](#) | [Visit American Memory](#)

### Library of Congress Online Catalog (at least 500 results)

Note: These results are sorted in alphabetical order.

1. [The 100th Congress /](#)  
Cahill, Nancy. | [1987] | text
2. [21st century prevention and management of migraine headaches : scientific program and abstracts : June 8 & 9, 2000, Hyatt Regency, Bethesda, Maryland /](#)  
National Institute of Neurological Disorders and Stroke (U.S.) | c2000. | text
3. [\[6 men working on construction of the attic octagon of the Thomas Jefferson Bldg. of the Library of Congress\] \[graphic\].](#)  
1892 Aug. 6. | still image

[See more results](#) | [Visit American Memory](#)

---

[About](#) | [Site Map](#) | [Contact](#) | [Accessibility](#) | [Legal](#)

# Metasearch

# Full Search Results

The LIBRARY of CONGRESS    [ASK A LIBRARIAN](#)    [DIGITAL COLLECTIONS](#)    [LIBRARY CATALOGS](#)

[The Library of Congress](#) > [New Search Home](#) > [Search Results](#)

[← Back to Results](#)

**Search Again**

**Select Sources:**

- U.S. historical, cultural collections
- Library of Congress Online Catalog
- Prints & Photograph Online Catalog
- Legislation, Congressional Activity
- Library of Congress Web site

[Search](#)

[→ New Search Home](#)

[→ FAQs](#)

[→ Search Help](#)

[→ Visit American Memory](#)

500 results from **U.S. historical, cultural collections**

Results 1 - 20 Page [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#)    [← Previous](#) | [Next →](#)

- [Silent Film Clip of Captain Pearl R. Nye](#)  
moving image
- [\[Thomas Jefferson, President of the United States\]](#)  
Tiebout, Cornelius, 1777-1830, engraver. | [1801(?)] | still image
- [Thomas Jefferson, third president of the United States](#)  
Pendleton's Lithography. | [1828(?)] | still image
- [1784 Convention between His Most Christian Majesty and the thirteen United States of North America : for the purpose of determining and fixing the functions and perogatives of their respective consuls, vice-consuls, agents and commissaries.](#)  
France. | 1788] | text

500 results from **U.S. historical, cultural collections**

Results 1 - 20 Page [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#)    [← Previous](#) | [Next →](#)

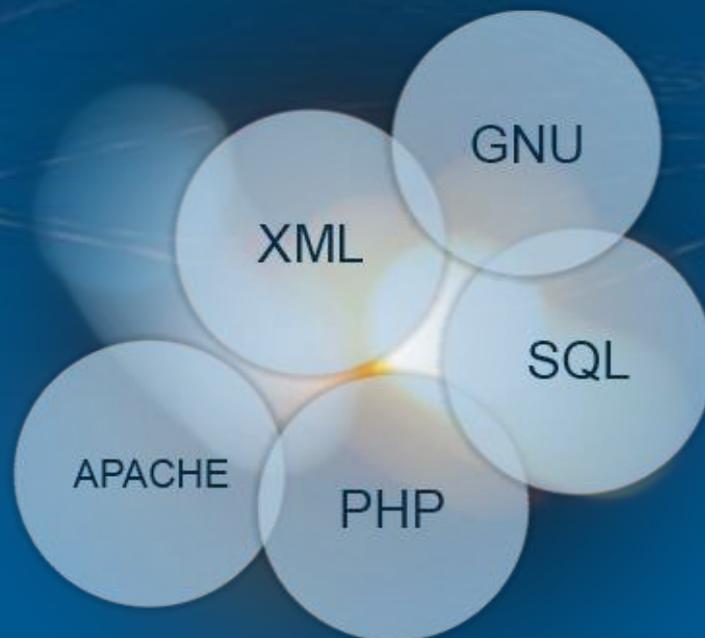
[About](#) | [Site Map](#) | [Contact](#) | [Accessibility](#) | [Legal](#)

# Metasearch

# Software Choice

## Index Data's Keystone Retriever:

- Open source (GNU General Public license)
- Protocol based (Z39.50, SRU/W)
- Previous experience with LC



<http://www.indexdata.dk/keystone/retriever/>

# Metasearch

# Keystone Functionality

## Ability to Search Remote Resources

- Query mapping on per target level in order to acquire meaningful results
- Communicates via Z39.50, extended to SRU/W search services via *simple2zoom*
- Gateways used for access to proprietary API's

# Metasearch | Interface

## Search Interface

- May range from simple search box to advanced
- Database selection options
- Complete customization of user interface: changes in formatting using CSS; content changes using XSLT

# Metasearch

# Search Results

## Processing of Search Results

- Grouping or merging of results
- Results all normalized to Dublin Core XML with XSLT and converted for display with XSLT
- Search results monitored and inserted in dynamic HTML using javascript and inline frames.

# Metasearch | Gateways

## Gateways to our 5 targets

- LC Catalog: Z39.50 access to LC Catalog already in place
- American Memory: gateway (Perl) provides SRU access to the homegrown application (InQuery search engine)
- Prints & Photographs: gateway (Perl) provides SRU access to the homegrown application (InQuery search engine)
- THOMAS: JAVA gateway provides SRU access to the application (K2 search engine)
- Web site: gateway implemented using Apache Cocoon to mimic SRU access to engine's built-in XML format (Ultraseek)



# Metasearch

# What Worked

- Improved discovery
- Worked collaboratively
- Experimented with a new model for change
- Increased understanding of underlying systems



# Metasearch

# Understanding Underlying Systems

- Highlighted need for consistent metadata
- Magnified differences in search systems requiring standardization

# Metasearch

# Feedback

- *“The LC’s new search is a commendable effort to wrangle simplicity and elegance from systems not on speaking terms with either.”*
- *“The new search tool helped me tremendously.”*
- *“I used it for the first time and it worked. Easy, simple and best of all it produced results.”*
- *“As a high school librarian, I LOVE your new federated search capabilities!!!! It has been too confusing for many of my students, who are required to use primary source documents in their research assignments, to effectively search through your wonderful collections -- until now!”*
- *“SUPER!”*

# Metasearch

# Next Steps

## Gradual rollout

- Implement on additional pages across site

## More targets

- Provide advanced option to search across many more properties

## Increased functionality

- Merged results sorted by relevancy rather than target
- Guided navigation
- Export, re-index catalog records to achieve relevancy ranking

# Metasearch

# Next Steps

## Gradual rollout

- Implement on additional pages across site

## More targets

- Provide advanced option to search across many more properties

## Increased functionality

- Merged results sorted by relevancy rather than target
- Guided navigation
- Export, re-index catalog records to achieve relevancy ranking

# Metasearch

# Improving External Access

## Open Archives Initiative

- "...develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content." (<http://www.openarchives.org>)
- Harvestable metadata in a variety of standard formats (dublin core, marc, mods) with embedded persistent URL linking to Library resource
- Representing close to 50 collections from American Memory and Prints and Photographs Collection
- Legacy conversion of data will make additional items available for harvesting

# Metasearch

# Sitemaps

- “Easy way for webmasters to inform search engines about pages on their site available for crawling.” “Sitemap 0.90 ... has wide adoption, including support from Google, Yahoo!, and Microsoft.” - [sitemaps.org](http://sitemaps.org)
- Used to expose “deep web” resources that are difficult to find via traditional crawling techniques.
- Helps overcome difficulties, e.g., lack of enumerated list of resources, temporary url patterns, and search-based navigation.

# Metasearch

# Sitemaps

- Example:

```
<urlset>
  <url>
    <loc>http://hdl.loc.gov/loc.pnp/cwpb.00981</loc>
  </url>
  <url>
    <loc>http://hdl.loc.gov/loc.pnp/cwpb.01030</loc>
  </url>
</urlset>
```

Metasearch

Contact Us

Chris Testa

[ctes@loc.gov](mailto:ctes@loc.gov)

David Woodward

[dwoo@loc.gov](mailto:dwoo@loc.gov)