



May 2008

# High Quality Discovery in a Web 2.0 World: Architectures for Next Generation Catalogs

John Mark Ockerbloom

*University of Pennsylvania*, [ockerblo@pobox.upenn.edu](mailto:ockerblo@pobox.upenn.edu)

Follow this and additional works at: [http://repository.upenn.edu/library\\_papers](http://repository.upenn.edu/library_papers)

 Part of the [Library and Information Science Commons](#)

---

## Recommended Citation

Mark Ockerbloom, J. (2008). High Quality Discovery in a Web 2.0 World: Architectures for Next Generation Catalogs. Retrieved from [http://repository.upenn.edu/library\\_papers/63](http://repository.upenn.edu/library_papers/63)

Presented at *PALINET Future of Cataloging Symposium*, May 2008.

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/library\\_papers/63](http://repository.upenn.edu/library_papers/63)  
For more information, please contact [libraryrepository@pobox.upenn.edu](mailto:libraryrepository@pobox.upenn.edu).

---

# High Quality Discovery in a Web 2.0 World: Architectures for Next Generation Catalogs

## **Abstract**

Issues of information and systems architecture underly many of the current debates over the future of cataloging. This talk discusses some ways in which the architecture of the catalog is being redesigned to combine the rich information architecture of library metadata with the robust systems architecture of many Web-based discovery systems. I will show "subject map" discovery systems that better exploit the relationships in complex ontologies like LCSH, and discuss a Digital Library Federation initiative to promote standards supporting interoperability between discovery systems and ILS data and services. I will also touch on the role of networked architectures in improving the quality and efficiency of library cataloging.

## **Keywords**

architecture, subject maps, interoperability, ILS-DI

## **Disciplines**

Library and Information Science

## **Comments**

Presented at *PALINET Future of Cataloging Symposium*, May 2008.

# High Quality Discovery in a Web 2.0 World

**Architectures for Next Generation Catalogs**

**John Mark Ockerbloom**

**PALINET Future of Cataloging Symposium**

**May 29, 2008**

# My talk: The one-slide version

- **“Web 2.0”, library catalogs have complementary strengths, weaknesses**
  - Library information strengths risk being left behind
- **The catalog needs to be re-architected, locally and globally**
  - Combine rich library information architectures with powerful “Web 2.0” system and social architectures
  - Innovate, but also harness “installed base” where possible
- **Catalog professionals should play important roles in the new architecture**
  - Planning its redesign, adaptation, and growth
  - Describing and managing a much larger network of cataloged resources, with rich information

# Architectures to consider (and examples I'll show)

- **Information architectures**
  - Example design: Subject maps for catalogs
- **System architectures**
  - Example design: ILS Discovery Interfaces
- **Social architectures**
  - Example design: PennTags

# Information architectures



# Some information architecture principles from “Web 2.0”

- **Design information structures for use**
- **Make simple information easy to use and express**
- **Make complex information possible to use and express**
- **Harness scale and complexity instead of fighting it**
- **Exploit all available information, resources, expertise**
- **Avoid unnecessary dependencies on transient technologies**

# Alphabetic architecture



(Photo by Mark Lindner, 2006. CC license: BY-NC-SA)



# Alphabetic catalog views

Franklin: Penn Library Catalog

http://www.franklin.library.upenn.edu/cgi-bin/Pwebi

Registry of Digital M... The Online Books ... Authorized Heading, ...

UPenn

Franklin: Penn Li... Franklin: Penn Libr... Amazon.com: wom... women - Google B... The Online Books P...

**Penn Library** | *Franklin: Penn Library Catalog*

new search search history saved searches my account place request bookbag user guide

Search Request: Subject = english language dialects england I  
 Search Results: Displaying 1 through 50 of 50 entries.

← previous next →

#	Titles	Headings	Headings Type
1	1	English language--Dialects--England--Lake District	Library of Congress
2	2	English language--Dialects--England--Lancashire	Library of Congress
3	1	English language--Dialects--England--Lancashire--Texts.	Library of Congress
4	1	English language--Dialects--England--Leeds.	Library of Congress
5	1	English language--Dialects--England--Leicestershire.	Library of Congress
6	4	English language--Dialects--England--Lincolnshire.	Library of Congress
7	1	English language--Dialects--England--Lindsey.	Library of Congress
8	22	English language--Dialects--England--London	Library of Congress
9	1	English language--Dialects--England--London--Texts.	Library of Congress
10	1	English language--Dialects--England--Lorton (Cumberland)	Library of Congress
11	5	English language--Dialects--England--Maps.	Library of Congress
12	1	English language--Dialects--England--Newcastle upon Tyne (Tyne and Wear)--Social aspects.	Library of Congress
13	3	English language--Dialects--England--Norfolk	Library of Congress
14	1	English language--Dialects--England, North--Maps.	Library of Congress
15	1	English language--Dialects--England--Northamptonshire.	Library of Congress

Find: elat Find Next Find Previous Highlight all Match case

Done

# Faceted architecture



(Photo by Romanlily, 2007. CC license: BY-NC-ND)

# Faceted catalog views

**Franklin**

Search Collection:

---

**Results**

Showing hits **1 - 21** out of **21**   Jump to: [First](#) • [Previous](#) • [Next](#) • [Last](#)

<p><a href="#">A study of Macbeth for the stage.</a></p> <p><a href="#">Accepted addresses, or, Proemium poetarum. To which are added, Macbeth travestie, in three acts, and miscellanies /</a></p> <p><a href="#">Selected poems /</a></p> <p><a href="#">Macbeth</a></p> <p><a href="#">Shakespeare's Lady Macbeth :</a></p> <p><a href="#">The tragedy of Macbeth /</a></p> <p><a href="#">Understanding Shakespeare :</a></p> <p><a href="#">Shakespeare's dramas of tyranny</a></p> <p><a href="#">Verdi's Macbeth :</a></p> <p><a href="#">Macbeth /</a></p> <p><a href="#">A textual study of Macbeth</a></p> <p><a href="#">Short studies of Shakespeare's plots /</a></p> <p><a href="#">Umanità e forza collate di Shakespeare :</a></p>	<p>Neilson, Francis, 1867- 1952</p> <p>1813</p> <p>MacBeth, George 2002</p> <p>Shakespeare, William, 1564-1616 1865</p> <p>Gerwig, George 1929</p> <p>William, b. 1867</p> <p>Shakespeare, William, 1564-1616 1965</p> <p>Frieman, Ruth 1964</p> <p>McGrail, Mary Ann 1988</p> <p>1984</p> <p>Shakespeare, William, 1564-1616 1971</p> <p>Amneus, Daniel J 1953</p> <p>Ransome, Cyril, 1851-1897 1924</p> <p>Sabbadini, Ado 1898 1954</p>
--	---

**Currently Used Filters:**

✘ Query : macbeth

---

**Narrow Results By:**

**Language**

<a href="#">eng</a>	18
<a href="#">ger</a>	3
<a href="#">fre</a>	1
<a href="#">ita</a>	1

---

**Subject** **all**

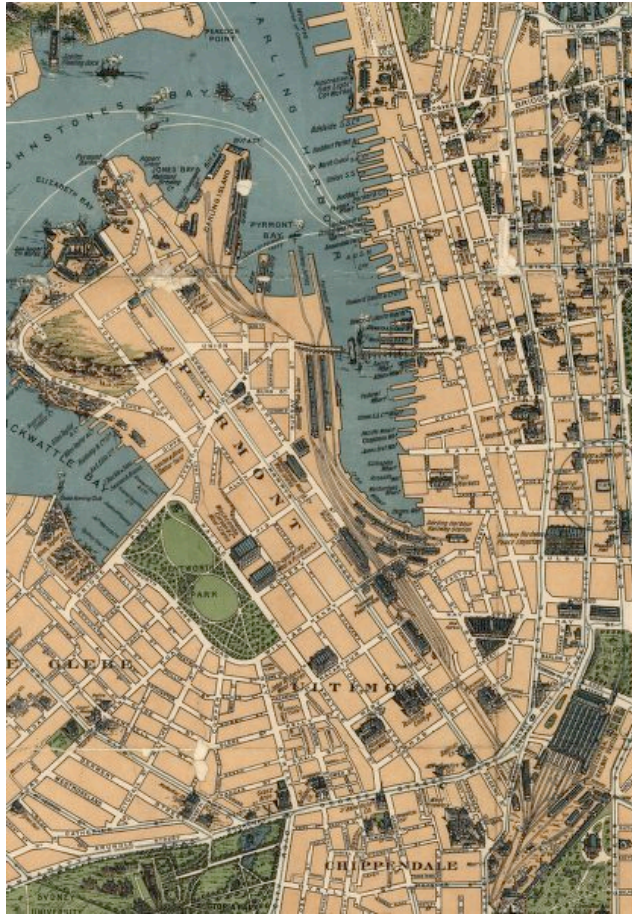
<a href="#">Shakespeare, William, 1564-1616. Macbeth.</a>	3
<a href="#">Shakespeare, William, 1564-1616 Tragedies.</a>	2
<a href="#">Abbaye de Saint-Wandrille.</a>	1
<a href="#">CHR 1825-1833</a>	1

---

**Name** **all**

<a href="#">Shakespeare, William, 1564-1616</a>	5
---	---

# Map-based architecture



From a 1922 map of Sydney,  
digitized by Library of Congress.  
Public domain.

# Map-based catalog views



Franklin Discovery - from the Penn Library Catalog

BETA



Browsing subject area: **Cataloging, Cooperative** ([About this browser](#))  
You can also [browse an alphabetical list](#) from this subject or from:

## Cataloging, Cooperative

### Broader terms:

- [Centralized processing \(Libraries\)](#)
- [Intellectual cooperation](#)
- [Cataloging](#)

### Related terms:

- [Bibliography, International](#)
- [Bibliography, Universal](#)
- [Library cooperation](#)

### Narrower terms:

- [Cataloging, Cooperative -- Bibliography](#)
- [Cataloging, Cooperative -- Data processing](#)
- [Cataloging, Cooperative -- Germany \(West\)](#)
- [Cataloging, Cooperative -- United States](#)
- [Cataloging in publication](#)

## Filed under: [Cataloging, Cooperative](#)

Full Title	Author	Date
<a href="#">British Library and AACR: report of a study commissioned by the Department of Education and Science; director of study A. H. Chaplin.</a>		1973
<a href="#">International distribution of catalogue cards, present situation and future prospects / [by] R. S. Giljarevskij.</a>	<a href="#">Giliĭ, ai ,revskiiĭ, R. S. (Rudzhero Sergeevich)</a>	1969
<a href="#">Locating books for interlibrary loan, with a bibliography of printed aids which show location of books in American libraries, by Constance M. Winchell.</a>	<a href="#">Winchell, Constance Mabel, 1896-</a>	1930
<a href="#">Goethe und die Weimarer Bibliothek, von Otto Lerche.</a>	<a href="#">Lerche, Otto, 1885-</a>	1929
<a href="#">Fuñnfzehn jahre Koñnigliche und staatsbibliothek, dem scheidenden generaldirektor exz. Adolf von Harnack zum 31. mañrz 1921 uñberreicht von den wissenschaftlichen beamten der</a>		1921

# How do you make the best subject maps?

- **Reuse what you can**
  - Relationships from LC authorities (just the start)
  - Subject assignments in existing catalog records
- **Automate what you can**
  - Subdivision, geographic, lexical, co-location analysis
  - Analysis can also automatically correct, localize subject headings
- **Specialize and refine where it gives the greatest benefit**
  - Logs can tell you what people are looking for, finding
  - You know what your special collections and communities are
- **Customize where you have to**
  - (but try to automate it, or share your customizations, wherever possible)
  
- **Had to open up the catalog system to do this...**

# System architectures

# Some system architecture principles from “Web 2.0”

- **Use the data for all it’s worth**
  - Analyze it, aggregate it, let it flow between systems
- **Encourage interoperation**
  - Standard formats, profiles allow data to be repurposed
  - Standard interfaces let lots of people invent new tools to interact with your information
- **Exploit the network**
  - Gives you access to more resources and smarts than you can draw on by yourself



# ILS Discovery Interfaces

- **Basic idea: Let any application use the data and services of your library**
- **Recommends standard functions all ILS's should support, gives a roadmap for implementation**
  - **Categories: Data aggregation, real time search, patron info and services, OPAC interaction**
- **Progress:**
  - **Digital Library Federation called task force last summer**
  - **Representation from 8 libraries (including LC, NLM, UC...)**
  - **Draft recommendation out (comment period just finished)**
  - **Official recommendation will be released in about a week**

# ILS-DI design principles

- **Distinguish abstract service, concrete binding**
  - Service: What the function should provide (semantics)
  - Binding: How the function should provide it (technology)
- **Multiple levels of interoperability**
  - From Level 1 (Basic discovery interfaces) to Level 4 (robust discovery platform that could replace an ILS's OPAC)
  - We pay particular attention to Level 1, and made detailed binding recommendations for it
- **Get requirements from libraries, commitments from developers**
  - Most ILS vendors agreed to provide Level 1 interoperability (in the “Berkeley Accord”)
  - We also encourage development by non-vendors
- **Quick and simple recommendations to support rapid prototyping, iterative development**
  - Level 1 already implemented for The Online Books Page
  - After the report, a workshop to promote development efforts

# Level 1: Basic Discovery Interfaces

- **Get bibliographic data out so it can be indexed and searched:**
  - Functions: HarvestBibliographicRecords; HarvestExtendedRecords
  - Recommended binding: OAI-PMH
- **Let users see what they can get now**
  - Function: GetAvailability
  - Recommended binding: REST/HTTP with XML response
- **Let users request them**
  - Behavior: GoToBibliographicRequestPage
  - Recommended binding: URL template (possibly OpenURL)

# A simple GetAvailability call

## Request:

[http://onlinebooks.library.upenn.edu/webbin/availability?id=olbp42044&id\\_type=bib](http://onlinebooks.library.upenn.edu/webbin/availability?id=olbp42044&id_type=bib)

## Response:

```
<dlf:collection
  xsi:schemaLocation="http://onlinebooks.library.upenn.edu/schemas/dlf/1.0/
  http://onlinebooks.library.upenn.edu/schemas/dlfexpanded.xsd">
  <dlf:record>
    <dlf:bibliographic id="olbp42044"/>
      <dlf:simpleavailability>
        <dlf:identifier>olbp42044</dlf:identifier>
        <dlf:availabilitystatus>available</dlf:availabilitystatus>
        <dlf:availabilitymsg>HTML at loc.gov</dlf:availabilitymsg>
      </dlf:simpleavailability>
    </dlf:record>
  </dlf:collection>
```

# What other standard interfaces could the catalog have?

- **Cataloging application interfaces?**
  - Automated quality control, subject assignment and checking, authority and subject map maintenance....
- **Item management application interfaces?**
  - Importing records from ERMS, publisher databases...
- **Collaborative cataloging interfaces?**
  - Data exchange with external cataloging partners?
  - Collaborative FRBR, authority management?
  - Integrating relevant non-librarian discovery data?
- **Collaboration implies social organization...**

# Social architectures



# Some social architecture principles from “Web 2.0”

- Encourage information sharing
- Encourage information repurposing
- Design incentives to contribute
- Design to scale up (resources and labor)
- Accept and adapt messiness

# PennTags: Sharing our finds

PennTags / search / discovery

***Everything is miscellaneous : the power of the new digital disorder / David Weinberger.*** copy

Weinberger, David, 1950- . *Everything is miscellaneous : the power of the new digital disorder / David Weinberger.* 1st ed. 9780805080438 series New York : Times Books, 2007.

Call#: Van Pelt Library HD30.2 .W4516 2007


Call#: Van Pelt Library HD30.2 .W4516 2007

tagged discovery by ockerblo ...on 21-MAR-08

***Garden State Discovery Museum*** copy

tagged libment kids by bmarcell ...on 07-MAR-08

***Chemistry: Chemical Synthesis, Drug Discovery, Materials Science, and Stable Isotopes*** copy



 subscribe to this page

search

**related to discovery**

- 2 + [dlf\\_spring\\_2006](#)
- 1 + [federated\\_search](#)
- 1 + [licensed\\_content](#)
- 1 + [recommender](#)
- 1 + [searching](#)
- 1 + [sru](#)
- 1 + [strategic\\_planning](#)

**some recent projects**

-  [Test Project](#)
-  [Alternative and Ethnic](#)



# Spreading PennTags around

**Location:** Van Pelt Library

**Call Number:** HD30.2 .W4516 2007

**Status:** Checked Out. Overdue on 03-12-08 (Use BorrowDirect+ or Place Request to Recall)

**Location:** Van Pelt Library

**Call Number:** HD30.2 .W4516 2007

**Status:** Checked Out. Due 09-30-08 (Use BorrowDirect+ or Place Request to recall).

[← previous](#)      [next →](#)

Print/Save/Email	
Select Download Format	Full Record <input type="button" value="Print/Save"/> <input type="button" value="Save Search Query"/>
Enter your email address:	<input type="text"/> <input type="button" value="Email"/>
<input type="button" value="Save"/>	<input type="button" value="Save To Bookbag"/>

[Proxy](#) | [Add to PennTags](#) | [Franklin Logout](#)

***This resource has been posted in PennTags***

***Posted by ockerblo on 03-21-2008***

*- annotation:*

*- tags: discovery*

# Social coordinators

- **Most useful shared resources have someone coordinating its development**
  - Can be active (e.g. Linus Torvalds with Linux)
  - Or passive (e.g. Penn Library with PennTags repository)
- **Must accommodate adequate scale, variety**
  - Both the information and system architectures important
- **Examples in library world**
  - Structures: MARC vs. FRBR/RDA...
  - Coordinator: LC vs. OCLC vs. LibraryThing vs. OpenLibrary vs. Google...

# The cost of locking data up

- **“Web 2.0 ... [is] really about data and who owns and controls, or gives the best access to, a class of data.”**
  - Tim O’Reilly
- **“Closed access is harmful to chemical data. That’s a fact, not a political stance. We are 10+ years behind other data-rich sciences because we protect data in archaic silos.”**
  - Peter Murray-Rust
- **“You should think of free as in free speech, not as in free beer”**
  - Free Software Definition

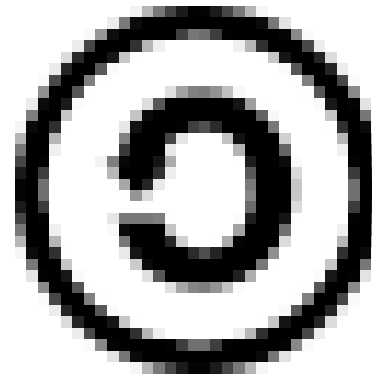
# Sharing catalog data

It could be public domain

Or: useful Creative Commons licensing for catalog data:



Attribution (BY)



Share-Alike (SA)

Not so useful for catalog data: Noncommercial (NC),  
No derivatives (ND)

# Accept and adapt messiness

- **Mess can be tolerated:**
  - Catalogs already have a lot of messy data
  - New techniques, tools help (auto-correction, fuzzy matching...)
- **Mess can tell us something useful:**
  - Tagging tells us how “real people” classify, find things
  - We can augment our subject taxonomies accordingly
- **Mess lets us scale up:**
  - Wikipedia lets lots more people build an encyclopedia
- **Mess can be progressively improved:**
  - OBP: From automated subject assignment to curated subjects
  - Penn videos: From hastily cataloged entries to detailed, high quality descriptions
  - Improvement targeted based on community needs

# Enhanced records in our video catalog

**VCat**  
Video Catalog

**Search**  
chintown  
All of these  
 DVD only  
GO  
Search hints...

[Advanced search...](#)

> **What's new in VCat?**

> **BROWSE BY:**  
Title  
People  
- Actors  
- Directors  
Genre  
Country  
Language  
Production Company

15 record(s) found for 'chintown' all word(s)

***Amarilly of Clothes-line Alley*** (1918)

Available. Van Pelt Video Collection; ask at Circulation Desk. DVD • 1999  
Call No. DVD PS3525.A489 A42 1999  
"Amarilly is the belle of Clothes-line Alley, a neighborhood near San Francisco's Chinatown. One night after a fight breaks out in the club where she works as a cigarette girl, Amarilly brings an injured socialite home with her. Grateful for her care, he hires Amarilly to clean his apartment and over time begins falling in love with her. But when Amarilly is presented at a posh social gathering, the disastrous results fuel the battle of high and low society."--Container.

***At the jazz band ball : early hot jazz, song and dance 1925-1933*** (1993)

Available. Van Pelt Library Ormandy Music and Media Center. VHS • 1993  
Call No. VHS Music Videorecording 475

***Chinatown*** (1974)

Circulation Review. Van Pelt Video Collection. VHS • 1999  
Call No. VHS PN1997 .C4644 1999  
Los Angeles private eye Jake Gittes (Nicholson) is approached by a mysterious woman (Dunaway) who wishes to have her husband investigated. As the case unfolds, though, he discovers it to be much more complex and dangerous than he had expected, involving politics, powerful men, and terrible family secrets.

***Chinatown*** (1974)

Available. Shelved at Storage. [Click here to request delivery.](#) VHS • 1999  
Call No. VHS PN1997 .C4644 1999  
Los Angeles private eye Jake Gittes (Nicholson) is approached by a mysterious woman (Dunaway) who wishes to have her husband investigated. As the case unfolds,

# Summary: Exploiting Web 2.0 design principles for discovery

- **General architectural recommendations:**
  - Make scale your friend
  - Free your data for sharing
  - Accept and adapt messiness
- **Use robust information architectures**
  - Your data should be thoroughly exploitable to the last byte
  - Harness new technologies to help exploit data (but realize that good catalog data outlives particular technologies)
- **Use open, scalable systems architectures**
  - Design for interoperation
  - Use the network to multiply your capacity
- **Harness social architectures**
  - Attract, coordinate communities to improve data and systems
  - Reuse, build on shared work; avoid redundant local work
  - Make the most of your own expertise and your communities'

# Continuing the conversation

- **Slides for this presentation**
  - [http://works.bepress.com/john\\_mark\\_ockerbloom/6/](http://works.bepress.com/john_mark_ockerbloom/6/)
- **Subject maps**
  - <http://labs.library.upenn.edu/subjectmaps/>
- **ILS Discovery Interfaces**
  - <https://project.library.upenn.edu/confluence/display/ilsapi>
- **PennTags**
  - <http://tags.library.upenn.edu/>
- **Mark Ockerbloom, John, 1966-**
  - Blog: <http://everybodyslibraries.com/>
  - Email: [ockerblo@pobox.upenn.edu](mailto:ockerblo@pobox.upenn.edu)
- **Let's talk!**

**These slides (except where noted) are CC-licensed: BY-SA**