

21 November 2008

# Today's Tabloid

PERSONAL NEWS FOR [rss4lib@gmail.com](mailto:rss4lib@gmail.com)

RSS4LIB

## RSS for Kindle Readers

NOV 20, 2008 10:48AM

Reading RSS on Amazon's Kindle reader through the web browser can be slow and inefficient. There's a new tool, [kindlefeeder.com](http://kindlefeeder.com) that automates the process by sending you, by email, the RSS feeds you wish to read. Once you've created an account at [kindlefeeder.com](http://kindlefeeder.com) and entered your Kindle user name, you then enter the RSS feeds that you wish to have mailed to you. The feeds are then gathered, converted into a Kindle-friendly format, and sent to your Kindle email account on your handheld device. The author of this tool, Daniel Choi, also built an open-source tool called [kindle-feeds](http://kindle-feeds) that allows anyone to set up a similar feed service for their own Kindle.



RSS4LIB

## Mashing Up and Remixing the Library Website — Access 2008

NOV 20, 2008 10:48AM

Karen Coombs

University of Houston A couple definitions: Mashing up — taking data from different places and shmooshing it together. Remixing — rethinking the way we do the library web site. University of Houston library web site, 3 or so years ago, had 1500 pages of content all managed centrally. Needed re-architecting for the same reasons we have all experienced: Staff have a wide range of HTML and computer savvy — running the whole gamut from technology illiterate to programmers. Library web sites incorporate information from lots of different sources. There's lots of redundancy of data — the same information appears on many different pages. Library users don't come (or wish to come) to library site — there's a need to get library information where patrons are. And finally, library information is not well integrated into the curriculum. The ACRL information literacy report reinforced that library instruction needs to take place as part of the curriculum, not as an add on. So — in the classroom, in the Course Management System. The traditional source has been to implement database-driven sources, skinning pages to look like each other. What Karen wanted was an easy to use system with little Web Services (her department) intervention. Content that can be easily mixed, reused, and shared within and between systems. Took inspiration from iGoogle, Drupal, netvibes, and WordPress. Liked content types, widgets (easy drag-and-drop customization). In the system built there, content owners are responsible

for content organization and metadata. Librarians have responsibility; supervisors have review role. A librarian owns pages but also items (building blocks of pages). Use many tools: LibraryFind, WorldCat, Archon, Serials Solutions, flickr, WordPress multiuser. The site is completely modular — librarians can add modules to pages, organize them how they want. Fonts and colors are controlled centrally, but layout is up to the content owner. Content is remixable. This means that content can be used elsewhere inside the site, but also used outside the site. The external uses aren't quite ready — the API is still in development. But they do use microformats, so if you view an event, the event is stored as an HCalendar microformat. Contacts are HCards, to be added directly to visitor's address book. Flash objects can also be embedded into a web page.

### Virtual Subject Library

This is essentially a subject guide. It brings together federated search, new books, relevant databases, subject liaisons, text blocks, etc. Sample (may not be permanent — this link is on a test server). Karen walked us through a quick demonstration of building a page. It's efficient and elegant — assuming that all the resources have already been added. An advantage to this is that there are no broken links. Or if there are, they are found and corrected very quickly since there is only one database entry. Most staff built their libraries in the course of a week — but it took more time for them to conceptualize *what* they wanted to have on the page and where to put it.

### Content Creation and Management

Tab editor works similarly. Through the admin interface, a librarian provides the names of the tabs, gives each tab a type, sets the order, and then (for each tab) makes it do the right thing. A search tab is configured to search something. A text tab is configured to display text. And so forth. An interesting add-on is that the link tool in the basic text editor takes the librarian to a search interface to all the links in the database. So an existing link can be added (and is therefore recorded once, used multiple times). New links can be added if the link is not already in the database. Since the site is modular, it is easy to replace functionality. Google Calendar could replace the current events tool. One staff directory could supplant another.

### Next Steps

It needs an API — badly. Both to integrate content into external sites and to improve internal AJAX content which will run much faster via API than via direct database query. While images and video can be uploaded easily, it's not easy to get them into a text block, for example — so this part is only half done. Integration with WorldCat — put bibliographies

from WorldCat into a page. Staff interface is very personalized. User interface is not, yet. A main impediment is that university information systems do not offer easy ways to get data about students (role, major, etc.) Bring federated search results into context of web site.

## Questions

**Q:** How does this work with accessibility?

**A:** The site, as seen by the public, is very accessible. The staff interface is not accessibility compliant yet — nor is it cross-browser compliant (Firefox only). **Q:** Can you talk about use cases for the API under development, and demand?

**A:** The first use case is internal users — to make current actions based on database connections will be faster via an API. Also, want to be able move things between Intranet and public site — some content belongs in both places. Desire to put library content into University web site. Courseware is difficult because UH is a WebCT site and that's not particular API-friendly. An API would make building applications for mobile devices easier. **Q:** Could you talk about development effort?

**A:** Been in the works for about 3 years. It's written in Cold Fusion. This was because old site was in that, and they didn't want the site frozen. They could replace parts as they went along. One developer FTE for one year built it; they now have two developers working on it.



RSS4LIB

# Using WorldCat Grid Services in Library Applications — Access 2008

NOV 20, 2008 10:48AM

Roy Tennant

OCLC Research Grid Services — a set of APIs to access WorldCat data. It's not for human consumption, but for machine-to-machine communication. We'll talk about a few services, with a few demos. Many of APIs — OCLC's and other's — are collected at TechEssence. These APIs are available for free to all OCLC member institutions.

## Identifier Services (xISBN, xISSN)

**xISBN** If you have a book ISBN, OCLC will return a list of all the related works (in a FRBR sense) to that book. It's a way to find a different edition, a different version, of the same book that might be in your catalog. Particularly helpful if someone comes into your catalog via Amazon or another bookseller — such as via the LibX Toolbar. **xISSN** If you give this tool an ISSN, it sends back a chart showing the history of that ISSN in all its splits, merges, renamings, and everything else that has happened. Example.

## Registry Services (Institution Registry)

Any institution can enter information for itself in this registry. It includes all sorts of things — hours, administrative contacts, root URLs for link resolvers, OPAC, and so forth. The WorldCat Registry also has an API.

## Experimental Services (Terminologies, Metadata Crosswalk)

**Terminology** Search for terms (broader, narrower, related) in various taxonomies: FAST, LCSH, MeSH, etc.).

## WorldCat Search API

This is the flagship OCLC API, released in August 2008 after an 8-month beta/test period. More than 80 institutions are signed up to use it. There are 110 million records, representing 1.3 billion holdings. This API supports OpenSearch and SRU. Responses come back in flavors of XML: RSS, Atom, MARC21 XML, and Dublin Core. JSON may be coming soon. It's RESTful. Many indexes. Sorts by relevance, author, title, date, libraries that hold it. It can return standard citations (APA, Chicago, Turabian, MLA, Harvard).



RSS4LIB

# We Love Open Source Software. No, You Can't Have Our Code — Access 2008

NOV 20, 2008 10:48AM

Dale Askey

Kansas State University

Libraries are not particularly good at making their own code open source and sharing. This is especially true of the small, lightweight applications that we build to make ILS systems “work right” or to solve small problems. These are frequently small (a few hundred lines of code). Why not? Several reasons... **Perfectionism** — the code's not ready yet, there are bugs, not commented well, it's inefficient. Even though it gets the job done — which is really what counts, many developers are hesitant to share their code with others. **Dependency** — we don't want to be supporting you; we can barely support ourselves. Putting it in a repository, with documentation — a good idea. Puts a bit of distance between developer (library) and user. Rutgers is planning to launch a library open source platform — but it hasn't happened yet (announced in April 2008). **Quirkiness** — What we do is so unique there's no point to releasing it; our problems are ours alone. This is false; while the exact problem may be different, but the general problem very often is shared. But — if you don't share the code, you end up with the full support and updating burden. There's nobody else who can help you find and fix bugs, add new features. **Redundancy** — Perfectly good software already exists that works for most people, so why should we offer our own? Good enough — the available solution — is often seen as better than doing it

oneself. **Competitiveness** — Our code is better than someone else's, so we want pride of ownership and don't want to share it. We build our own to be the best, not to share the technology. Institutional Repositories are a case in point — institutions develop/implement their own but all too rarely share their successes to save others time. **Misunderstanding** — Administrators do not understand nature of OSS tools — they understand and know how to deal with vendors. Functionality can be built on a good foundation — the open source tool — and customized. This is the antithesis of what vendors offer. Open source puts responsibility for getting it right in institution's hands, not in a vendor relationship.

### What Can We Do?

Figure out a way to share software among libraries. There are methods for "big stuff" (Koha, Evergreen, etc.). But what about small stuff? Several initiatives, but none global Google code is one, but it doesn't meet everyone's need and isn't accessible to non-technical librarians. A library-specific repository might be useful. Put a license on our code and let it go when asked to share it. Even for the small snippets. Commit to the necessary human investment to build and maintain open source software for our own good. Reward staff for contributing to open source communities. This should be viewed as a form of professional development/contribution. Re-prioritize internally to make open source contributions happen. Favorite soundbite from Dale's talk: "Minesweeper is like digital heroin."



RSS4LIB

## Thunder Talks — Access 2008

NOV 20, 2008 10:48AM

Thunder Talks are brief (4 minutes 30 seconds!) talks on any subject the speaker wants to talk about. Without further ado:

### BiblioCommons

Rolled out at Oakville (Ontario) Public Library. Live implementation of the tool (the research leading to BiblioCommons was described in one of yesterday's post. Lesson learned: Don't roll out a JavaScript-heavy site on IE 6 browsers. In a few weeks, have received thousands of ratings, reviews, and lists.

### OLE — Open Library

OLE is a framework for libraries that support research, teaching, and higher education. Led by Duke University. By July 2009 there will be a completed design document. Focuses on design, not software development, but expects that a follow-on implementation phase will happen.

### How To Adopt Open Source

Things they've done recently in University of Prince Edward Island library catalog. Since there was no reserves module in Evergreen, they used book bags to gather resources and use RSS to populate subject guides. Added linkable subject terms. Switched thumbnail images from Amazon to Google Books API. Also added a tab in the catalog record for "Excerpt", which pulls an excerpt of the book from Google Books.

### What's Happening in Saskatchewan

One system for all public libraries to share resources. Single library card, no ILL — just place a hold. This project was going well, and then it wasn't when the budget went away. But progress was still being made. Libraries created consortia, did an RFP, and are moving forward. A consortial integrated library system (CILS). Start of a three-year process.

### RSS Feeds for New Books

Based on Doran's New Books List for Voyager catalog. A 3 element table: Use a single LC class, a multi-character LC class (A, B, and J), or ranges of call numbers (AB 123 to AB 155). Script outputs both RSS and HTML. They can be mixed and match to pull together a new books list that matches specific patron needs. See KSU's New Books Feed. Custom feeds are possible — for KSU staff only — but requires that patron enters request in words, library staff translate into LCSH speak. But custom feeds are underutilized. Only 10 custom feeds, mostly for librarians, so far.

### Koha in a Small Public Library

Hanover, Ontario, public library wanted to go in with a small number of libraries (14) nearby. Couldn't find a Koha support vendor to do the implementation the way they wanted. So they're doing it themselves. Building on virtual machines to allow flexibility — in implementation, to add new libraries.

### Fedora Drupal Module

University of PEI built a Fedora module for Drupal. Will be open-sourced "very soon." Have several content models. Makes ingest very easy — to solve the basic problem of many repositories: nobody contributes because it's hard. Have an example of a RefWorks collection — to ingest citations and, if allowed locally, if it's allowed to put in the full text (if ROMEO allows it). Metadata are editable in Drupal but stored in Fedora.

### Zotero Connection to Evergreen

If you have the Zotero Firefox plugin, you see Zotero cues in the Firefox URL bar. It then lets you select items from the results list into Zotero. Accomplished through a LINK item in the document head that points to the unAPI service that Zotero uses.

### Dashboard for Library Information

Summary information for administrators. It shows trends — not

individual items, but gross numbers over time that can help administrators understand what is going on. There's a back end that allows users to easily create widgets, pulling together numbers from various systems. The tool allows easy creation of these widgets, but the data are entered by hand. A storehouse for library data with graphical presentation.

### Drupal Module for ContentDM

A Drupal module built at Simon Fraser University to search ContentDM through Drupal for the Multicultural Canada site. It uses the ContentDM API and the Drupal API, or "one big 'appy family."



RSS4LIB

## MyLibrary: A Digital Library Framework and Toolbox — Access 2008

NOV 20, 2008 10:48AM

Eric Lease Morgan  
University of Notre Dame

MyLibrary is about creating relationships. It's a way to catalog resources — very broadly defined (people, databases, books, you name it). MyLibrary invented about 10 years ago, had a lot of success/popularity then. Concept of "my library" picked up by others such as MyLibrary, etc. It was a turnkey application — download, install, and run. It was simple, and it worked, but wasn't as complex. MyLibrary is made up of four kinds of resources:

1. Resources
  2. Patrons
  3. Librarians
  4. Facets and terms
3. All of these resources are stored in "Dublin Core-esque" data structures. Patrons in system have name, major, etc. Librarians have name, subject areas, contact info, etc. Resources have material types, subjects, academic level of primary audience. All of these descriptive terms are "facets and terms". Facets are classes of terms. For example: format: book; subject: forestry; and so on. You can have as many facets as you like, and as many subjects under each facet. It's all 2 levels deep.

### 3. Examples

3. Examples from the Notre Dame site:

- Research tools — lists of research tools
- Subject
- Reading list — combination of things classed "format: journals" organized by subjects. This was created via OAI from the Directory of Open Access Journals. Specific subjects or specific journals can be added to "my library."
- Facebook MyLibrary widget. It's not "facebook" that's important. The fact that the MyLibrary toolbox allows it to happen is important.
- FAQs — each frequently asked question/answer pair is a resource. They're cataloged. Then they are browsable and can be displayed on relevant subject or topic pages.

3. MyLibrary is not meant to do everything — just managed "piles of stuff". It does not support search. It does not support OAI. But data can be pulled out of MyLibrary and fed to a search engine. For example, Alex Catalogue of Electronic Texts. Browsable and searchable lists of 14,000 full-text public-domain books. MyLibrary is not a particularly strong open source project — there isn't a community around it, for which Eric takes blame. It's in Perl, but that's a passé language now. Coming up is a web services interface on top of it, probably Atom. But some sort of RESTful web service is coming. Question: It's been in operation about 5 years; how are students using it?
3. Answer: Students don't know they're using it. They don't customize it — it's just the way the web site work. Question: What are privacy issues with patron data?
3. Answer: Librarians take privacy more seriously than patrons. Patrons expect easy to use interface that gets them what they want. Libraries are behind the curve on this. MyLibrary makes some broad guesses about what patrons are likely to want. Any future personalization effort will be opt-in. Individuals won't get assigned resources, but aggregates (freshmen, math majors, etc., not John Smith).



RSS4LIB

## User-Generated Content and Social Discovery in the Academic Library Catalogue: Findings from User Research — Access 2008

NOV 20, 2008 10:48AM

Martha Whitehead, Queens University

Steve Toub, BiblioCommons

The problem is “discovery” — getting answers to questions that you don’t know how to ask. In other words, finding things you don’t know about. Not just updating the catalog. They were dissatisfied with the federated search tools. Catalogs are solitary experiences, but learning and research are social activities. User-generated content is what this project is about. Narrowly, tags, ratings and reviews. In the broader sense, curating that information. The research project with BiblioCommons was aimed to figure out how tagging works in the academic environment. Reading lists are an obvious, and old, form of user-generated content. Research paths in libraries — how to do subject research — are another (librarian-generated, but we’re users, too). Faculty members are the “ultimate research advisor.” The ideal research process, in an Ontario Council of University Libraries study, users want to see recommendations from “authorities,” wanted to find classics in the field, and also wanted to find surprises — serendipity. Draws a distinction between social discovery and social networking. The former is serious. What features should be built into an academic research site? Fear that information would be misleading, that faculty (who know subjects best) wouldn’t have time to contribute, that students (for any number of reasons) won’t contribute. But students are inherently social and even when in the library want friends to know where they are. Study participants wanted to know what their trusted colleagues (professors, fellow researchers) think.

### User Research in Academic Environment

BiblioCommons is a next generation discovery tool, a social network, and an OPAC. In March 2008, Steve Toub recruited Queens University faculty, students and librarians to talk about how they do their research. Non-librarians do not limit (i.e., use facets) very much. Students don’t reformulate queries; they go back to original search and re-do it. Users would avoid LCSH at all costs in the catalog (but would use it as a browsing tool). Students don’t “experience pain” when manually formatting citations — it’s just part of the process. Librarians think direct export to RefWorks a must. Librarians want to help; users want to be independent. Second round of research in June about user-generated content (UGC). Went through a variety of tests, from paper prototypes to full mock-ups. Focus of this study on a very narrow sense of UGE. Not much understanding of why people should tag. Not clear understanding of motivations for tagging in the academic library catalog. Started by asking students, if you want to buy a camera or see a movie, where do you go? Asked if students looked at comments by others. Most of this 18-22 age range said they sought out sites with UGE. Preferred comments from “people like them” over recognized critics/reviewers/professionals. Most had used ‘Rate My Professors.’ They mostly looked at comments, not paying lots of attention to ratings. One student said, “I don’t necessarily want the opinion of a professor — I’m looking for people who are as incompetent as I am.” Two of 10 students knew what tagging was by name. But they didn’t have any idea what it was when they saw the MTagger tag cloud. Tagging in Facebook pictures is utterly different from tagging text. Ideas to change it: change labeling — use “themes”, “keywords”, “what terms you use to help others find this”, explain tagging in the cloud, not via a link. In the catalog, showed a mockup of a review system. For recently-returned books, user can say how useful it

was and for which course (from list). Provide brief survey of what user used in the book (whole thing, just a chapter and which one), etc. Another version — provide sliders for “relevance to course”, “level of difficulty”, “personal interest”, etc. The most important data element requests were things like “is it going to be on the test”? “How is this related to other texts”? “How is this related to the lecture”? Users requested clear signals about how important the item is to the class. Most students wouldn’t fill in more than 1-2 data elements — so opportunity to collect data is limited. Most wanted anonymity, one that’s not personally identifiable. Most students wanted to share their comments — that was the point. Asked, what if — when you logged in to the library — you saw the syllabi for your courses? Very popular. When are students most likely to contribute? Only if syllabus is online. Probably not for current week’s syllabus, for immediately previous week. Netflix-style “you just returned this item, would you rate it?” sort of interaction. Putting collection point for UGC at the right point in the workflow is trick. What kinds of rewards are of interest? From list of choices, top response was to help others get to resources faster. Idea of “paying it forward” — if I do it now, it will help others later, which will help me when I need it. Sense of “empty restaurant syndrome” — if no tags are there, why would I join in? “Buying” student participation seems pretty easy. Barriers to contributing: nobody wants to support freeloaders (help those who don’t contribute), but I want to have content there when I want it. Fears of plagiarism overweigh willingness to share with others — even at level of sharing reading list for a paper through an online system. Three strategies for ensuring quality: 1) Authentication — people log in to library and library knows who you are (even if it’s not your university ID). 2) Aggregation — pool content from multiple systems provides more content and helps “smooth out” details. Ability to identify individual users while seeing the mass. 3) Marketplace of ideas — create a self-managed system (no editorial review) to make sure reviews themselves are vetted by the masses. BiblioCommons roadmap Near term — provide an outstanding user experience — make interface simpler, cleaner, and more intuitive. Mid term — organize catalog experience around courses and assignments — not LCSH or broad subject guides. You see a course- or assignment-specific view when you log in to the catalog. Long term — breaking down barriers between silos. Federated search isn’t the answer. Everything is integrated. BiblioCommons Status User research led to current priorities. This year and next — an iterative beta release process.



RSS4LIB

## Drupal: Content Management and Community for your Library — Access 2008

NOV 20, 2008 10:48AM

Ilana Kingsley  
Dave Mitchell  
Harish Nayak

Debra Riley-Huff  
Nick Ruest

### University of Alaska, Fairbanks (Ilana Kingsley)

**Movie collection:** Movie covers (built with Drupal 4). Pulls in movie records from catalog (it's a Sirisi catalog, so they need to screen-scrape) and matches with images and ratings from the Internet Movie Database and/or Rotten Tomatoes. **Library web site:** Ilana got tired of making small changes to site and wanted to get staff more involved in content editing. Using Drupal's modules, can customize what appears where and when. Looked at lots of CMS tools (leaving out Plone, since Ilana didn't know Python). Installation was easy, didn't need to know lots of PHP. There's a huge Drupal community — lots of support. Had a two-year implementation process. Part of problem was political; campus IT department was not in favor of PHP/MySQL. Content analysis was a key element — making sure she understood the content types so that, ultimately, they could all be defined in the database and then assigned to individuals for maintenance and upkeep. Keeps updating/adding modules — after testing on a development server. Has a number of content types: Advertisement, annual report, collection guides, exhibits and collections, news & events, article indexes and collections, etc. Roles form basis for content types. Roles started with departments.

### University of Mississippi (Debra Riley-Huff)

**Subject guides:** Used content construction kit to create a content type for subject guides. Customized navigation and presentation. The Presidential Debate guide (set up for the first U.S. Presidential debate (at the campus) got heavy use. The Drupal install held up well under heavy traffic. Themes are what makes a Drupal site look like you want it to. You can make Drupal content look the way you want it to. Best to start with "Zen" theme, which is bare-bones and easier to customize than out-of-the-box themes that come with Drupal. Matching existing site is difficult. Relied heavily on Content Construction Kit. **Government documents:** A government documents repository site — government documents librarian can maintain the content through Drupal.

### University of Rochester, River Campus (Harish Nayak)

Revamped library web site into Drupal. Also, Drupal is being used in the eXtensible Catalog (XC) project at Rochester, so there's a large internal drive to make it happen there. Their redesign process involves numerous activities: Several items center around the content: User research — the library has a staff anthropologist at UR did an ethnographic survey of how students use the library (broadly, not just online). Technology — showing new technologies to library staff. Usability — this is the checkpoint to make sure that the technology is being applied in good ways. Design — where the programming requirements come from. These are all interconnected in various ways. Customization of user content (through MySite and/or Panels themes) gives a more personalized user experience. Rochester used MySite to allow users to rearrange their pages. Relies on JavaScript in the page. More interaction with server is necessary (pages aren't all the same for all users) so can increase load.

### London Public Library (Dave Mitchell)

Picked Drupal because of cost. But got very easy customization as a result. Modified the comment tool so that comments could exist across sets of pages, not just on a single page — so that, for example, election information comments and questions could appear on all government-related pages as a single thread.

### Nick Ruest (McMaster Library)

Library's Digital Collections. Drupal isn't an out-of-the-box digital collections tool, but Drupal's CCK allows for the creation of Dublin Core metadata set. OAI-PMH & CCK: The site has been harvested by several OAI-compliant harvesters, putting digital content into broader access.



RSS4LIB

## Open++: Dispatches from the OSS Frontlines — Access 2008

NOV 20, 2008 10:48AM

Keynote: Karen Schneider  
Community Librarian  
Equinox Software (Evergreen)

Karen Schneider on "Open++: Dispatches from the OSS Frontlines". Karen's job is to travel around Georgia talking with libraries around the state helping them with the Evergreen installation.

### Evergreen

We've seen lots of open source software in libraries in recent years. Tons of experimentation has taken place, lots of it by and for libraries. Pines had a need for a consortial catalog for 270+ libraries across a large state. Some vendors said they couldn't do it; others offered far-too-expensive options. In 2004, a development team started building their own ILS. In 2006, Evergreen was launched in 200 libraries. Version 1.4 is imminent (first couple weeks of October). They have kept a tight development cycle. Key point: With Evergreen, librarians are once again writing their own ILS. This is analogous to what happened about 30 years ago (for example, with the Melvyl catalog, built at home in pre-vendor times). In recent decades, libraries strayed from path of doing their own stuff and went down the vendor path. Now, we've come full circle and libraries are once again starting to take control of their own destiny again. Network effect has been huge (combined with general state of economy and price of fuel): holds and interlibrary loans within the Evergreen system is growing exponentially. There are now 275 libraries in PINES. Other consortia include those in British Columbia (Sitka) and Michigan. There's also an academic installation — Indiana University — live now. In development are other consortia, including an academic one. But these are just the "known" sites — it's open source, so many other K-12 schools likely use it.

## Observations about open source and libraries

1. Documentation is critical — must be a formal requirement. Documentation doesn't come easy. Evergreen got a Mellon grant to write it — but that's not the normal path.
2. Trickle-up Engagement: Originally, it was thought that libraries would automatically know how to “do open source.” However, that turns out not to be the case. Libraries need some help getting started — getting re-engaged in the software development process.
3. Gift economy: Community around Evergreen is small, skilled, and dedicated — a smaller community of developers than they initially expected. People contribute actively, though not as broadly.
4. A surprising revelation: end-users are all alike - but library workflows are unique. Users are much more similar than libraries. Evergreen has a very flexible back end. This turns out to have been a very good idea. Flexibility in the workflows is critical.

## Features of Openness

Open has several positive features. Communication becomes distributed — no longer vendor-contact, it's many people looking, many people fixing. Many eyes makes a better product, with many hands to fix them. The network effect is significant on the library side: the more libraries participate, the better. Local issues and requests lead to global improvements. Customization is the user side of back-end flexibility. Fosters partnerships — there's no need for secrecy, keeping vendors in the dark about local implementation, and libraries in the dark about vendor plans. Cost — it's not necessarily cheaper to go open source, but it moves the costs around from licensing to updating/maintenance.



RSS4LIB

# VuFind: The Library OPAC Meets Web 2.0 — Access 2008

NOV 20, 2008 10:48AM

Andrew Nagy  
VuFind Developer  
Villanova

## Introduction

What is a “next generation catalog”? The term is not Andrew's favorite — he wants to get away from the word “catalog” and start talking about “resource discovery.” A different approach between librarians and users: we view things as known-item searches, users have no idea, generally speaking. We need a tool to facilitate browsing, sharing, and organizing resources. Users go to Amazon, Google, and Delicious to find and discover, and then look in the catalog. Libraries should be in that

discovery role. Definition of “catalog” should include things that your users have access to (as through consortial borrowing or online access). Villanova decided to turn product into an open source one because they wanted broader development base and broader help for making it better. It took about 2 months to get university approval to open source the software. Over next two years, development continued at VU and elsewhere. Many institutions are in process of adopting VuFind — alpha, beta, and live. VU's catalog is tightly integrated with the web site. Browsing is important, along with functions that exist in Amazon, Barnes & Noble, elsewhere — both online and in physical stores. Bring in demographics — show sophomores, for example, what other sophomores have looked at in a given topic search. Ability to save and share searches is also important. Villanova found that, in the physical library, students were confused by multiple service points (reference, information, circulation, etc.). They've combined the physical desks into one — and feel very strongly that students would do better with a single point of service via the web. Catalog, web, digital collections, search — should all be integrated. Tool must integrate directly with the ILS — LDAP & SIP2 authentication, bring in live circulation status, display holdings data, and — most importantly — interoperate with major ILS systems. Data migration — VuFind community built SolorMARC to import MARC data into SOLR. You specify the mappings of MARC records into SOLR, the way you want. Investigating OAI Import and (possibly) Z39.50 import.

## Search and Browse

A VuFind search (see the VuFind demo) for a phrase like *world war two* (no quotes) does an “and” search first, and then an “or”, so it searches for all the words, then any of the words. For phrase searching, use quotes. You can narrow results using the facets VuFind returns. Author facet for a general search is not particularly helpful — for a broad search, some authors always show up (Shakespeare shows up in the *world war two* example, and they clearly were not writing about World War II. Nagy is thinking of ways to have facets display in a different order depending on the kind of search being done. Suggests removing author from “very large” systems (those with millions of records). An author search in the demo brings up a list of matching authors before the search results — so users can disambiguate. Clicking on an author's name from this list pulls up the Wikipedia entry for that author, so that users can verify which of the similarly-named authors they meant. There's a experimental browse tool that lets you navigate through the catalog, iTunes-style, without typing a keystroke, to get to a collection of books on a topic. This tool only shows top 50 results in the last panel — so it leaves out a lot of things, even in a smallish catalog like the test catalog (with 850,000 records).

## Questions

**Q:** What are weighting systems in the search process?

**A:** In general, the all fields search gives more weight to title, exact matches, author, call number, subject headings. It's not currently configurable, but it's in the code and you can play with it there. **Q:**

Where is VuFind in development process for bringing in non-catalog materials?

**A:** Villanova has a lot of digital library material. But haven't brought in

non-MARC records. Want search results to show thumbnail of digitized object. A record display would include different types of data. This is probably the next step. VuFind 1.0 is next step. Bringing in other content through OAI and/or federated search is likely the next step. **Q:** How frequently is bibliographic content?

**A:** Villanova updates/adds/edits/deletes about 150 records a day, on average — they update nightly. It can be done more often, if wanted/needed. Their Voyager ILS does a nightly update of the previous day's changed records. Deletes and suppressed records are separately output and removed from VuFind nightly. **Q:** Have there been any III libraries?

**A:** None public, but quite a few are in development. Holdings are currently available only through screen scraping. Sirsi-Dynix is also under development. **Q:** Internationalization?

**A:** Yes — the whole interface has been translated into about 10 languages.




---

RSS4LIB

## Tentative Settlement in the Google Book Search Lawsuit

NOV 20, 2008 10:48AM

The path is now open for millions of books digitized through the Google Book Search project to be available to the public. This includes in-copyright books as well as out-of-print and out-of-copyright books. The tentative agreement announced today would settle the class action lawsuit filed against Google in 2005 by the Authors Guild, the Association of American Publishers and a handful of individual authors and publishers. The big news, from the perspective of libraries, is that libraries will be able to provide their patrons with access to vastly increased numbers of digital volumes. (See the Google Book Search Copyright Settlement for the full text and details.) Two points are worth noting in particular:

- 1. Free online viewing of books at U.S. public and university libraries** — “public libraries, community colleges, and universities across the U.S. will be able to provide free full-text reading to books housed in great libraries of the world like Stanford, California, Wisconsin-Madison and Michigan [the Google Book partners]. Public libraries will be eligible to receive one free Public Access Service license for a computer located on-site at each of their library buildings in the United States. Non-profit, higher education institutions will be eligible to receive free Public Access Service licenses for on-site computers, the exact number of which will depend on the number of students enrolled.”
- 2. Institutional subscriptions to millions of additional books** — “[L]ibraries will also be able to purchase an institutional subscription to millions of books covered by the settlement agreement. Once purchased, this subscription will allow a library to offer its patrons access to the incredible collections of Google’s

library partners from any computer authorized by the library.” In other words, libraries can add Google Books to their proxy servers (for a fee) and thereby allow full-text access to millions of books.

While the settlement has not been approved by the courts, the news that Google and the publishers have agreed on terms is terrifically exciting. Disclaimer: My employer is one of the Google Book partner institutions.




---

RSS4LIB

## Liability Insurance for Bloggers

NOV 20, 2008 10:48AM

With the rise of blogging as a recognized form of journalism “has come greater scrutiny and the inevitable rise in legal threats facing bloggers,” says David Cox of the Media Bloggers Association (MBA), a not-for-profit, non-partisan organization supporting the development of blogging and citizen journalism. The MBA has recently announced a program to offer “liability insurance program for bloggers which provides coverage for all forms of defamation, invasion of privacy and copyright infringement or similar allegations arising out of blogging activities.” As bloggers take up larger roles in journalism, public commentary, and social discourse, the individuals and organization they write about are increasingly paying attention. The risks of being accused of libelous, defamatory, or other language are the same as in any other media; the world is now paying more attention. The MBA now offers an online course, “Online Media Law: The Basics for Bloggers and Other Online Publishers,” without charge. Upon completing the course, students are offered the opportunity to join the MBA and then to purchase (at a discount) the liability insurance. Anyone who has taken the course has access to directories of attorneys specializing in online libel cases. Is it worth the cost of an MBA membership (\$25/year) and insurance (I could not find details of the insurance cost on the MBA site) to mitigate against what I suspect is a small risk for me? Probably not, in my case. The more controversial a blogger’s posts, though, the more likely it is that someone might find them legally troublesome (and not just annoying).

### Related Post

*RSS and Legal Liability (4/24/2008) Disclaimer: I have no affiliation whatsoever with the Media Bloggers Association.*

Via RSS Blog.





RSS4LIB

## Bloglines Update

NOV 20, 2008 10:48AM

There's been a lot of discussion (see TechCrunch, , and Law.Librarians, among others) about Bloglines and the problems they've been having. I noticed today that Bloglines has returned to the *status quo ante* — RSS4Lib's subscriber numbers, as inaccurate and wonky as they may be, have returned to where they were before I noticed the precipitous drop. Looking at my log files, I see that the Bloglines crawler now reports that I have 799 subscribers to the RSS feed (and 56 to the Atom). That actually represents growth over the last consistent numbers Bloglines provided via the crawler and, not unusually, a few more than Bloglines reports on its web site (its Beta site is more up-to-date, reporting the same 799 RSS feed subscribers as its crawler does). Bloglines posted on its technical blog yesterday a brief note about the outage. It was Apple-esque in the level of details — it offered none — other than to say that the problem was fixed:

Some folks might have noticed that specific feeds were not updating recently on Bloglines, and we wanted to update you and fill you in on what's been going on. We have figured out what the glitch has been. Over the weekend, a fix was released on Bloglines to resolve the issue. All feeds should now be updating and back to normal. If you're still experiencing problems you can report a stuck feed.

I still prefer Bloglines to Google Reader (call me old-fashioned), but was about to make the leap. I'm pleased the Bloglines is still alive and keeping their crawlers and index going. Google needs the competition — even if it's not as serious as it could be.

RSS  
4Lib

RSS4LIB

## Did Bloglines Purge Its Subscription Rolls?

NOV 20, 2008 10:48AM

I just noticed that the number of subscribers at Bloglines recently fell sharply. (I use RSS4Lib Feedstats tool to track my blog's readership via RSS.) On September 27, Bloglines was reporting a total of 872 subscribers to the RSS and Atom feeds from my blog (see details). This number is consistent — Bloglines has reported a gradually growing number of readers, adding a few each week, over the past months. On September 28, there were only 565 subscribers (see details) to the two feeds, according to Bloglines. Bloglines reports the number of users who read each feed in the server log files. For example, the Bloglines crawler passed through earlier today and left this log line: 65.214.44.28 - - [06/Oct/2008:08:12:38 -0700] "GET /index.xml HTTP/1.1" 304 - "-" "Bloglines/3.1 (<http://www.bloglines.com>; 527 subscribers)" The number of Bloglines subscribers went down for both RSS and Atom

(from 796 to 527 and 56 to 38, respectively). These lower numbers have stayed consistent since 9/27/2008, which makes me think it's not just a transient error. Interestingly, the numbers reported for each feed within the Bloglines web site have not changed. (See Bloglines' list of RSS and Atom subscribers, neither of which has been updated.) The web site's numbers have always lagged the Bloglines crawler's numbers by a week or more, so the discrepancy itself is probably not significant. Has anyone else noticed that Bloglines subscriber numbers took a dive a couple weeks ago? Can anyone with a long-dormant Bloglines account confirm that it has been purged?

RSS  
4Lib

RSS4LIB

## RSS to iPhone / iPod touch

NOV 20, 2008 10:48AM

DoYouFeed, a site that takes an RSS feed and formats it for pretty viewing on an iPhone or iPod touch. Once you give it an RSS feed to process, it returns an iPhone- iPod touch-friendly web page with the headlines and brief descriptions from the feed. Clicking on an article link pulls up the full text of the item but without the blog's trappings — very handy if you're using your gadget on a slow data network (as opposed to wireless). While the reader itself works as advertised, I have two criticisms. The first is that the iPhone's version of Safari already knows how to read RSS feeds — it does so by routing them through a service that Apple provides. The second, and more serious one, is that DoYouFeed puts advertisements at the bottom of the full text article, thereby making money from the blogger's words without returning any of it to the author. In the case of the DoYouFeed version of RSS4Lib's feed, this is in direct violation of the Creative Commons Attribution-Noncommercial 3.0 license I placed in the RSS feed. This license only allows, as it says, non-commercial use of the feed. DoYouFeed should parse the RSS feed and either not display non-commercial-use only feeds or not display advertisements on them.

RSS  
4Lib