



Stanford University **LIBRARIES &
ACADEMIC INFORMATION RESOURCES**

Blacklight at Stanford: A Highly Leveraged, Reusable, Discovery Engine

Tom Cramer
Associate Director
Digital Library Systems & Services
December 2009

Blacklight at Stanford

Blacklight is an open source, "next generation" discovery application that works equally well for library catalogs as for digital repositories. Originally developed at the University of Virginia, it has since been adopted by a growing number of institutions, including Stanford University.

Within the Stanford Libraries' digital ecosystem, Blacklight is being leveraged as a central component in a broad range of applications, from a library "discovery layer" to a repository front-end, from a digital manuscript viewer to a catalogers' tool.

This is an overview of Blacklight, its capabilities and its advantages as a highly reusable discovery application. It outlines Stanford's rationale for choosing Blacklight, provides examples of the system in operation in a variety of contexts, and covers the state and direction of both the code and the community for this open source project.

Vast Information Resources

6,825,821

6+ million
monographic
and serial works

800,000
maps & images

25,000
E-journal
subscriptions

800
Licensed
databases

20
Different
libraries

1
Department of
Special
Collections

The Features You'd Expect

- **Faceted search**
- **Relevance ranked results**
- **Personalization (bookmarks, tags)**
- **Export citations via RSS, SMS, Email, Zotero, RefWorks, EndNote**
- **Sort by anything <date, author, relevance, title>**
- **Support for unicode / vernacular scripts**
- **And much more...**

Plus Three Key Features

- 1. More than MARC: support for any kind of record or metadata**
- 2. Object-specific behaviors**
 - Books, Images, Music, Manuscripts, Finding Aids, <any>
- 3. Tailored views for domain or discipline-specific materials**
 - UVa's "Music Tab"
 - UWisc's union catalog

Object Specific Behaviors & Non-MARC records

Limit Search by...

- Format
- Subject
- Call number
- Region
- Source
- Collection

Catalog Music

All Fields Results per page 10 Search

Displaying items 1 - 10 of 576 for no searches in items with SOURCE U.Va. Art Museum (Start over ?)

Sort by Date Received

Get updates with RSS

1 2 3 4 5 6 7 8

1 Lucius Verus, A.D. 162

Mint Rome
 Format Coin
 Location The University of Virginia Art Museum
 Accession 1989.19.14

2 Valentinian II, A.D. 388

Mint Trier
 Format Coin
 Location The University of Virginia Art Museum
 Accession 1989.19.23

Facets are tailored to numismatics

Note the "Source" facet is the U.Va Art Museum tab.

Search results data fields are customized to content type

from <http://blacklightdev.lib.virginia.edu>

Blacklight: Tailored Views Based on Context

The screenshot shows the Blacklight search interface. At the top, there are tabs for 'Catalog' and 'Music', with 'Music' selected. A search bar contains the text 'brahms'. Below the search bar, it says 'Displaying items 1 - 10 of 1,776'. A 'Sort by' dropdown is set to 'Relevancy'. On the left side, there is a 'Limit Search by...' section with various facet filters: Format, Recordings and scores, Recording format, Instrument, Composition era, Author, Library, Language, Source, Region, and Subject. The search results are displayed in a list format. The first result is 'Join' by Daniela Brahm, a book published in 2005. The second result is 'Brahms', a cassette recording from 198-?. The third result is another 'Brahms' recording by Biret, Idil. Red arrows point to the 'Music' tab and the search results, with text boxes explaining that the facets are tailored to music resources and that the search results are only for music.

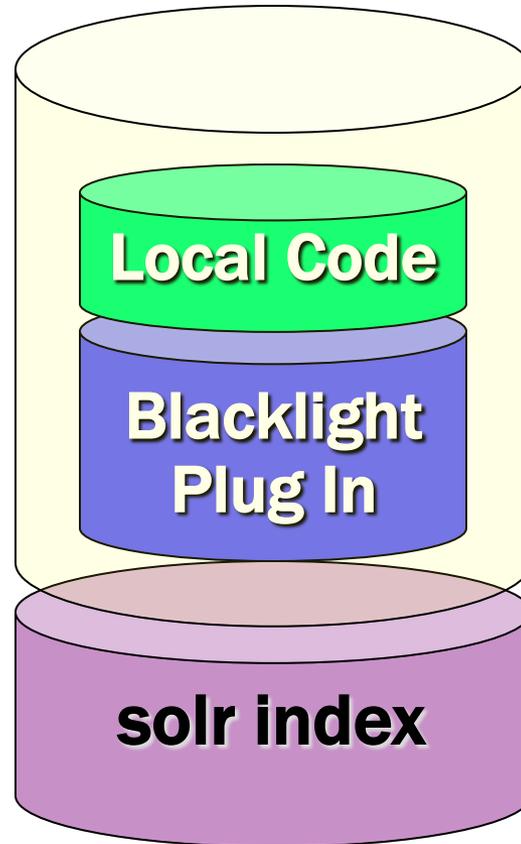
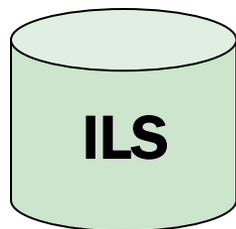
Facets are tailored to Music resources & discipline

Note the "Music" tab.

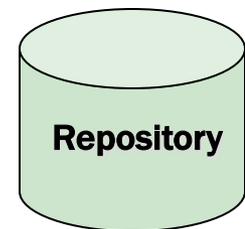
Search Results are only for Music

Blacklight's Technology Stack

Ruby on Rails
Application contains
both the *Blacklight
plugin* and *local code*.
Local code augments
and over-rides (where
needed) the BL plugin.



An underlying *Solr
index* holds data from
sources of interest.
Indexers prepare and
load data into usable
form.



Ruby on Rails

- **Rapid** application development for web applications: “Convention over configuration”
 - 10x productivity
- **Supportable**: MVC (Model-View-Controller) structure separates data from logic from presentation
- **Testable**: Rspec and Cucumber give powerful, automatable test coverage tools
- **Learnable**: Stanford went from 1 to 8 Ruby savvy developers in one year (no new hires)
 - 1 week learning curve to basic proficiency

Blacklight at Stanford

- Next generation catalog
- IR discovery (ETD's and more)
- Digital manuscript search
- Digital archival collection front end
- *Repository administrative interface (coming soon)*

Hydra: One Solution for Many Needs

Hydra is an effort that is developing and packaging an application framework to sit atop Fedora, and tailoring the use of this framework for specific institutional repository & digital library solutions. It is a joint development project among Stanford, University of Virginia, University of Hull and DuraSpace

The spirit of Hydra is a common body (framework), many heads (tailored UIs). The Hydra/Fedora hybrid will produce IR applications which support the following common functions:

- * Deposit
 - * Edit & Annotate
 - * Set Permissions / Access Levels
 - * Manage collections
 - * Report
 - * *Browse*
 - * *Search*
 - * *View Object*
- } Blacklight provides the search, browse & viewing capabilities

ETD Application

Search for student, advisor, title, department...

[Start over](#) | (Search/browse selections display here)

Browse ETD collection

Degree

[Masters](#) (5,505,751)
[PhD](#) (266,989)

School

[Business](#) (532,536)
[Earth Sciences](#) (836,431)
[Education](#) (532,636)
[Engineering](#) (5,505,751)
[Humanities & Sciences](#)
[Law](#) (532,536)
[Medicine](#) (590,959)

Status

[Submitted](#) (6,431)
[Rejected](#) (959)
[Verified](#) (2,536)
[Available](#) (32,536)
[Embargoed](#) (532)
[Withdrawn](#) (36)

1 - 10 of 12345 results sorted by **author**

[< Prev](#) [Next >](#)

Author/Title	Degree	Advisor
Alter, Susan Elizabeth Demographic and evolutionary processes in baleen whales inferred from ancient and modern DNA	PhD	Palum
Plumlee, Megan H Occurrence and photochemical fate of emerging contaminants: Nitrosamines and perfluorochemicals	Masters	Daily,
Shester, Geoffrey G Sustainability in small-scale fisheries: An analysis of ecosystem impacts, fishing behavior, and spatial management using participatory research methods	PhD	Loague
Brosi, Berry Justice Bee communities, pollination, and landscape context in tropical countryside	PhD	Daily,
Carr, Adrienne Elizabeth Physics-based simulations of hydrologic-response and cumulative watershed effects	PhD	Loague
Ward, Gillian Statistics in ecological modeling: presence-only data and boosted MARS	Masters	Haste
Jones, Nicole Louise The role of wind-waves and currents on vertical mixing in shallow water bodies: Implications for phytoplankton distribution	Masters	Monisthorm
Cornwell, Will Causes and consequences of functional trait diversity: Plant community assembly and leaf decomposition	Masters	Ackert
Weng, Kevin Chi-Ming	PhD	Block, Barbara A

Search & Browse powered by Blacklight

Search for student, advisor, title, department...

[Start over](#) | [Back to search results](#)

Actions

Submitted 2/22/2009 3:20pm

Rejected 2/24/2009 12:39pm
 Margin spacing is incorrect. Please reformat so margins are 1" on all sides, then resubmit.
 - John R. Doe

Resubmitted 2/26/2009 2:55pm

Next step

- [Reject](#) (return to student to resubmit)
- [Comment](#) (no change in status)
- [Verify](#) (sign off and send to library for final processing)

Your comment here...

#3 of 10 results for Degree: PhD, Status: Submitted

[< Prev](#) [Next >](#)

Name: [Shester, Geoffrey G.](#)
 Degree: PhD
 Major: School of Earth Sciences: Interdisciplinary Program in Environment and Resources (PER)
 Advisor: [Loague, Keith](#)
 Dissertation Title: Sustainability in small-scale fisheries: An analysis of ecosystem impacts, fishing behavior, and spatial management using participatory research methods

Thesis or Dissertation (PDF)

File Name	Size	Date
Shester-Sustainability-in-small-scale-fisheries.pdf	1,060K	2/26/2009 2:51 pm
cover.jpg	125K	2/26/2009 2:52 pm
video.avi	2,325K	2/26/2009 2:52 pm
data.xls	325K	2/26/2009 2:52 pm
illustration.jpg	600K	2/26/2009 2:52 pm

Abstract

Anthropogenic impacts to ocean ecosystems have been well documented throughout the globe, highlighting commercial fishing as a primary threat to marine life populations and habitats. Small-scale commercial fishing operations provide more than half of the world's seafood, employ over 90% of all fishermen, and take place along a vast proportion of the world's coastline. Their local, decentralized nature poses unique challenges for research, data collection, and management, particularly in developing nations. As a result, a focus on large, industrial fisheries has discounted the potential impacts of small-scale fisheries, leading to an assumption that they are generally more sustainable. My research on two remote fishing cooperatives in Baja California Sur, Mexico asks to what extent these small-scale fisheries are sustainable, and in cases where significant ecological impacts do occur, I offer and compare several distinct conservation approaches. While engaging fishermen to participate directly in the data collection, I begin by conducting a series of studies that examine how small-scale fisheries interact with marine ecosystems. I quantitatively assess potential ecological effects including bycatch of non-target species and damage to seafloor habitats resulting from several common fishing practices and find that the type and magnitude of impacts are significantly different across small-scale fishing gears. Set gillnets have higher bycatch and habitat impacts than other fishing gears, representing a major priority for small-scale fishery conservation. I also analyze fine-scale fishing behavior in the lobster fishery to help explain why some fishermen are more successful than others and how marine habitat characteristics and resource distribution influence their gear placement strategies. Observable trap movement behaviors on a limited number of fishing trips

ETD-specific viewing behavior

09 Available

Digital Medieval Manuscripts

blacklight

in All Fields Search

[Login](#) | [Search History](#)

Limit your search

Collection

[Parker Medieval Manuscripts](#)
(559) [\[remove\]](#)

[Goodspeed Manuscript Collection](#)
(0)

[Handschriften der Abtel Pfäfers](#)
(0)

[Oxford Medieval Manuscripts](#) (0)

Country

Institution

Material

[Vellum](#) (390)

[Paper](#) (127)

[Vellum and paper](#) (18)

[Paper and vellum](#) (9)

[Vellum \(and paper\)](#) (4)

[Paper \(and vellum\)](#) (3)

[Paper and Vellum](#) (2)

[vellum](#) (2)

[Paper \(1st and last leaves](#)
[vellum\)](#) (1)

[Paper \(and some vellum](#)
[documents\)](#) (1)

[Paper with vellum flyleaf](#) (1)

[Vellum fragments mounted on](#)

You searched for:

No Keywords

✓ [Collection](#) > [Parker Medieval Manuscripts](#) x

Displaying items **1 - 10** of **559**

[Start over](#)

[« Previous](#) [Next »](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) ... [55](#) [56](#)

Sort by relevance

Show 10 per page

1. 1

Collection: [Parker Medieval Manuscripts](#)

Manuscript Identifier: 1

Provenance: I can discover in the book no hint of its monastic provenance. The upper R. corner of ff. 1r-2v is mutilated: probably the press-mark was there.

2. 2I

Collection: [Parker Medieval Manuscripts](#)

Manuscript Identifier: 2I

Provenance: The book comes from Bury St Edmunds Abbey. On the upper corner of f. 2r (first leaf of text) is a mark rather smaller than in most Bury books, but of the same kind: B. 1. The press-mark shows that this was the first book in class B. The B here stands for Biblia. Further, at f. 322r the edge of the leaf has been mended with a patch of vellum in cent. xv on which is sketched a crowned head (cut off at the neck) and a scroll inscribed hic, hic, hic. This represents St Edmund's head, which called out Here, here, to those who were searching for it after the martyrdom. It fixes the provenance in a very satisfactory way. In the old catalogue (cent. xii, xiii) of the Abbey books preserved in a MS. at Pembroke College and printed in my *Essays on the Abbey of Bury*, 1895, p. 23, the second item is *Bibliotheca in duo uolumina* (!): and in the *Gesta Sacristarium Arnold Memorials of Bury St Edmunds Abbey* (*Rolls Series II*, p. 290) in the account of Hervey

Digital Medieval Manuscripts

Solr Admin (dmss)

sulwebappdev1.stanford.edu:8100

cwd=/home/dsgprod/app-solr/apache-tomcat-5.5.26 SolrHome=/opt/home/dsgprod/app-solr/work/dmss_solr/



Solr	[SCHEMA] [CONFIG] [ANALYSIS] [SCHEMA BROWSER] [STATISTICS] [INFO] [DISTRIBUTION] [PING] [LOGGING]
App server:	[JAVA PROPERTIES] [THREAD DUMP]

Make a Query [\[FULL INTERFACE\]](#)

Query String:

Assistance

[\[DOCUMENTATION\]](#) [\[ISSUE TRACKER\]](#) [\[SEND EMAIL\]](#)
[\[SOLR QUERY SYNTAX\]](#)

Current Time: Mon Dec 14 11:44:46 PST 2009

Server Start At: Thu Oct 22 13:48:34 PDT 2009

A solr index without the Blacklight frontend.

Multi-Institutional Project

- **Originated at UVa in 2007 as a research project**
 - Moved to production as “Virgo Beta” in 2008
- **Stanford adopted in Jan 2009**
 - Deployed SearchWorks on Blacklight in Aug '09
- **Currently dozens of installations**
- **~10 committers from 6 institutions (UVa, Stanford, Nat'l Library of Agriculture, Johns Hopkins, UWisc)**

Test Coverage

- **Full test coverage is a core community principle**
 - Unit tests with Rspec
 - Acceptance tests with Cucumber
 - Continuous integration testing with Hudson
- **Tests ensure...**
 - Quality
 - Compatibility
 - Clarity of code and function
 - Confidence

Testing is a Core Community Principle

See http://projectblacklight.org/?page_id=2

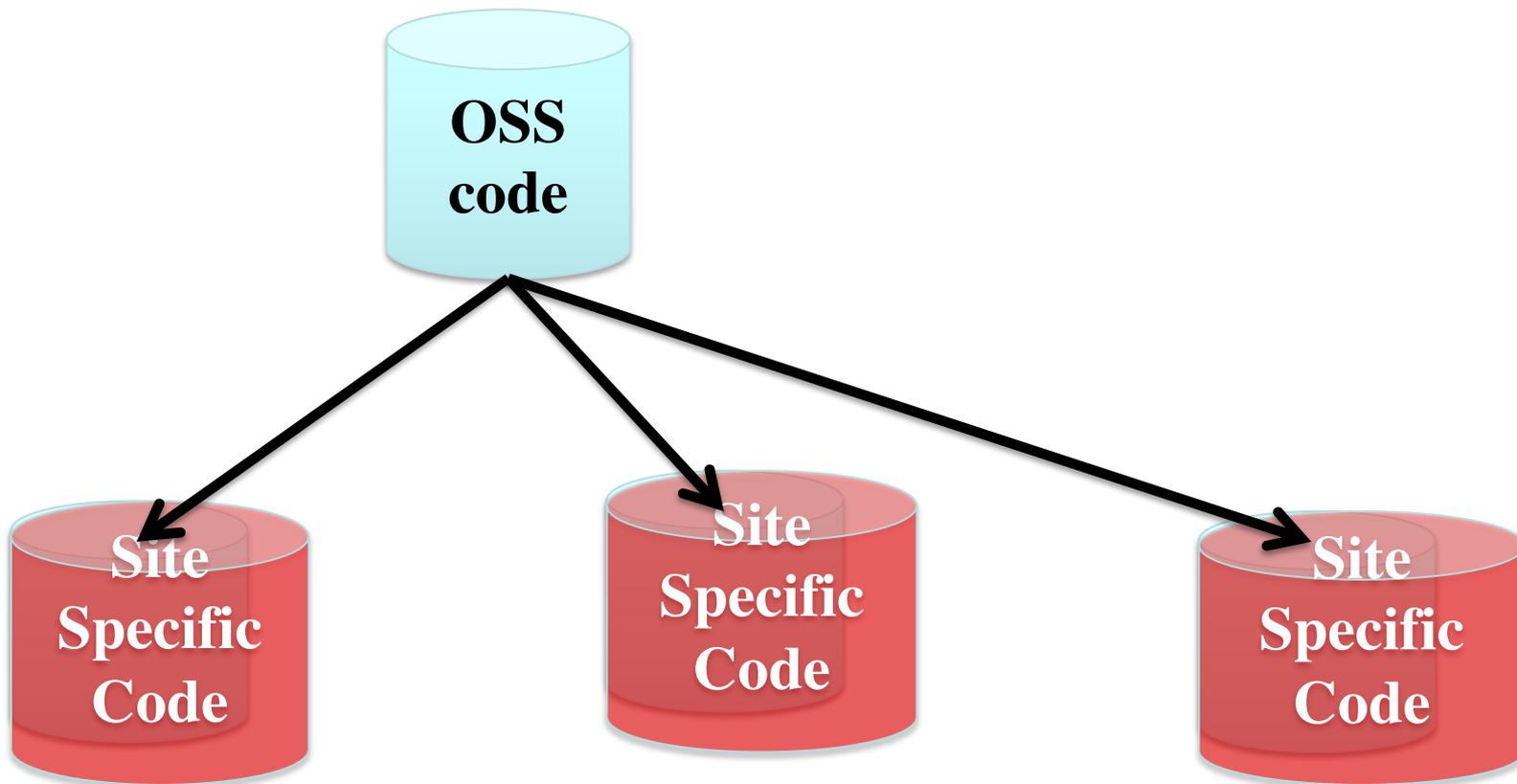
- ***“All contributed code must have full test coverage before it is committed. The current testing infrastructure is RSpec for everything but Rails views, and Cucumber for for Rails views.***
- ***“Tests must be committed at the same time code is.***
- **“All bugs and development tasks will be tracked in JIRA**
- **“All code must be documented before it’s committed.”**

BL's Current Test Coverage is 90%

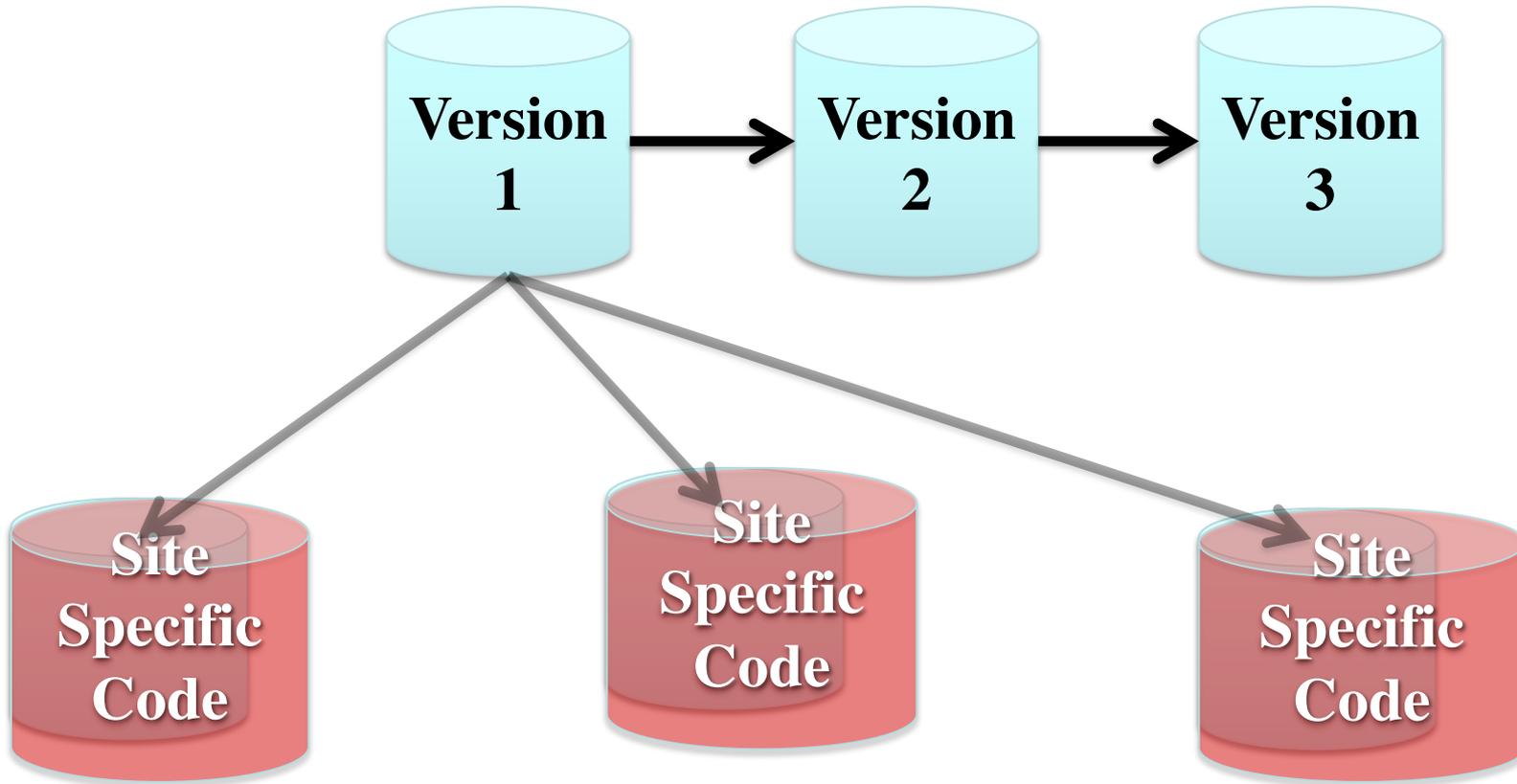
Name	Total lines	Lines of code	Total coverage
TOTAL	2073	1357	90.55%
app/controllers/application_controller.rb	59	59	96.61%
app/controllers/bookmarks_controller.rb	59	59	100.00%
app/controllers/catalog_controller.rb	200	137	96.00%
app/controllers/feedback_controller.rb	37	24	35.14%
app/controllers/saved_searches_controller.rb	44	34	79.55%
app/controllers/search_history_controller.rb	25	21	76.00%
app/controllers/user_sessions_controller.rb	43	20	100.00%
app/controllers/users_controller.rb	13	8	100.00%
app/helpers/application_helper.rb	354	246	90.96%
app/helpers/bookmarks_helper.rb	3	2	100.00%
app/helpers/catalog_helper.rb	44	29	90.91%
app/helpers/feedback_helper.rb	2	2	100.00%
app/helpers/saved_searches_helper.rb	2	2	100.00%

<http://hudson.projectblacklight.org/hudson/job/blacklight-plugin/99/rcov/>

The Code Silo Problem



The Code Silo Problem

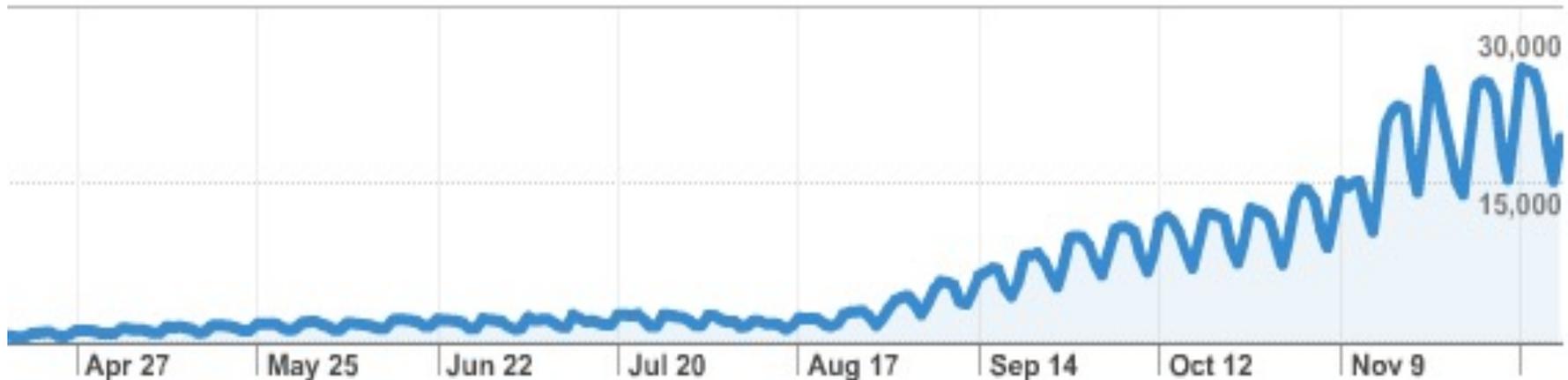


Well-Structured Code

- **Blacklight 2.0 was a substantial refactoring to make the code portable**
 - **Core functions, common to all installations, located in a plugin**
 - **Local modifications made in the Ruby on Rails application container**
 - **Over-rides facilitate customization for local needs**
- **Vendor drops are straightforward**
- **GIT to facilitate branching and merging**

Scalability: SearchWorks = known upper bound

SearchWorks Usage: April – December, 2009



- SearchWorks currently has > 6 Million records
- Peak daily load is now > 26,000 visitors

Why Blacklight?

- Reusable tool for both catalog, repository and digital library application front ends
- Ruby on Rails is an excellent platform choice for digital library development (velocity, supportability, testing)
- Code structure and engineering support distributed development and tailored uses
- UVa (et al.) are excellent community partners

Resources

<http://projectblacklight.org>