Hello. I’m George Oates, Project Lead of the Open Library project, from the Internet Archive in San Francisco.
Wow! Congratulations on 10 years! I’m Very Impressed by the strength of the Koha Community and the remarkable distributed development that’s built this awesome software. Congratulations on being one of the web’s best open source systems.
I’m a web designer by trade, so had a fairly steep learning curve into the world of libraries and bibliographic metadata when I started the job in the middle of last year. As I began to learn the lay of the land, Koha quickly became apparent as a first class piece of open source library software, and I wrote to Chris to see if I could come to KohaCon, and... here I am!

I’m bummed that I can’t stick around for the Hackfest later, so hopefully, I can show you enough about Open Library to get you interested to use it as a resource... But first, a short introduction to the Internet Archive...
I work at the Internet Archive, leading The Open Library project. We recently moved in to this church in The Richmond in San Francisco. We’re turning it into a library.
We’re based in San Francisco, California, where I happen to have been living for about 5 years.
It’s a great town, and if you ever come, let me know and I’ll take you out for a drink!
Since 1996, the non-profit Internet Archive has been building a digital library of Internet sites and other things in digital form. archive.org has a ton of texts, video, software, live music... all sorts of things.

Our mission is Universal Access to all Knowledge. Not a bad reason to get out of bed each day...
It’s not your traditional non-profit... Lots of the staff are technologists and developers.
We have many computers. They store over
- 100,000 hours of TV from channels all over the world
- 250,000 moving images or video
- 500,000 audio recordings
- 2.5 million scanned texts
- 150,000,000,000 web pages
Just the other day we had 2.88 petabytes of hard drives delivered. That’s enough storage for about 2 billion books.
One of the more well-known things the Archive produces is The Wayback Machine. It’s been capturing and storing the Internet since 1996. For example, this is a snapshot of the National Library of New Zealand’s website, captured on July 7, 2000.
Quite a change to today’s site...
Another service the Archive offers is something I thought the librarians amongst us might be interested in. It’s sort of like a mini Wayback Machine that people can use to curate web collections around a certain topic or event.
For example, this is an initiative of Columbia University Library's Center for Human Rights Documentation and Research, the Human Rights Web Archive is an effort to preserve and ensure access to freely available human rights resources created mainly by non-governmental organizations, national human rights institutions, and individuals.

Europe-focused human rights organization based in Germany. Includes reports. In English; some content in Danish and German.
  Ethiopia-focused human rights organization based in Ethiopia that promotes democracy and the rule of law. In English.
  Eritrea-focused web archive set up to compile all human rights violations committed by the government since 1991. The archive also provides documentation on atrocities committed against Eritrean civilians by the Ethiopian government between 1954-1991 and during the 1998 border war. Contains photos, video, and audio. In English.
  Southern Africa-focused human rights organization based in Johannesburg, South Africa, which works toward "the promotion of credible elections, participatory democracy, human rights culture, and the strengthening of governance institutions." Contains reports. In English.
  Great Britain’s national human rights institution. Includes publications. In English.
- [http://www.erichina.org/english/englishhome.htm](http://www.erichina.org/english/englishhome.htm) - Empowerment and Rights Institute
  China-focused human rights organization based in Beijing, which "focuses on protecting the human rights of disadvantaged groups including farmers, laborers (esp. female and child labor), migrant workers, complainants and petitioners, and victims of forced-relocation." Contains Reports. In English (Chinese version under construction).
  Guatemala-focused organization based in Guatemala that uses forensic anthropology to investigate and document human rights violations. Includes reports. In Spanish and English.
  Francophone Africa-focused women's rights web site based in Dakar, Senegal, which provides information in support of sustainable development and gender equality. Contains reports. In French.
  Chile-focused Christian human rights organization based in Chile. Contains reports. In Spanish.
  Africa-focused forum based in Kenya. Aims to to "promote girls' and women's education in sub-Saharan Africa in line with Education For All." In English.
- [http://www.fhri.or.ug/](http://www.fhri.or.ug/) - Foundation For Human Rights Initiative
  Uganda-focused human rights organization based in Kampala that is committed to the protection and promotion of human rights and democracy. Contains reports. In English.
- [http://www.flamnet.info/](http://www.flamnet.info/) - Forces de Libération Africaines de Mauritanie
  Mauritania-focused political organization based in Dakar and Paris that seeks the establishment of a new form of government in the country in which there is equality between the Black African minority and the Arabo-Berber majority. Contains multimedia elements. In French.
- [http://www.focusweb.org/](http://www.focusweb.org/) - Focus on the Global South
  Southeast Asia-focused anti-globalization organization based in the Thailand, the Philippines, and India. In English.
  Asia-focused human rights organization based in Geneva, Switzerland, that strives "to empower people by advocating social justice, sustainable human
Another major part of what we do is scanning books. This is a picture of one of the scanning centers in San Francisco. We currently employ about 200 staff scanning books.
And today, we have over million free texts available online - that includes over 1 million books
150 million pages scanned
1,000 books scanned EVERY day
24 scanning centers in 5 countries, and we hope for more.
We’re also scanning microfilm, which is much faster than individual books. Here’s an example of the record of the population census from 1790 to 1930. Scanned from microfilm from the collections of the [Allen County Public Library](http://www.allencountylib.org) and originally from the United States [National Archives Record Administration](https://www.archives.gov).
Dear Aunt,

In pursuance of a long-ago promise,

I send a few lines to express to you how much I have been settled down in the West. I have been busily engaged with my friends, and have lived in a sort of seclusion, but am now on the move. I am about to go to the mountains, and am looking forward to the pleasure of a visit to some of the wonderful natural scenes. I am expecting to return in a few weeks, and shall be glad to hear from you.

Yours sincerely,

[Signature]
Over 1 million free books that you can read on archive.org today, and access through the Open Library site, by checking the little “Only eBooks” box as you search.
As well as being able to download these books in a variety of different formats, from PDF to TXT and more, we also have a web-based book reader, which you can use to read our scanned texts within your web browser, without the need for any additional software.

I’ve actually come to Wellington direct from a meeting in San Francisco called Books in Browser, held at the Internet Archive last week. It was there that we announced an upcoming new release of our bookreader, which will hopefully go live in the next few weeks... Here are some screenshots...
The main reason we wanted to improve on the current design was to try to build an “app-level quality” book reading experience right in the browser. This included several improvement for touch interfaces in browsers on devices like the iPad.

From a straightforward design perspective, there were also improvements to be made on usability and simple stuff like making the book bigger in the browser window.
This is a screenshot with the toolbar open, where you can see new features like a navigation bar at the bottom that allows you to scroll through the book, a “read to me” feature which plays the book in a computer-y voice, and highlights what’s being read. Also, if we know a table of contents for the book, each chapter is mapped along the navigation bar.

We’ve also rewritten the full text search engine, and I’ll talk more about that a bit later.
Apologies for the slightly blurry picture, but this is my boss, Brewster Kahle, who founded the Internet Archive back in 1996. He’s playing with a touchscreen which is displaying the new bookreader. The screen’s been installed in one of the reading desks that used to sit in the reading room of the Christian Science church before it became our new home. A big part of the bookreader redesign was to evolve an app-level quality book reading experience within a web browser. If you have an iPad, I’d encourage you to try it!
Before I move on to talking about Open Library, I wanted to show you something specific that the scanning co-ordinator in Boston drew my attention to – it’s Isaac Newton’s OWN COPY of his book, Opticks. It has Newton’s own handwriting and annotations in the scan... SO COOL.

And now, I want to shift gears into talking about Open Library, designing a web-based catalog, and direct Koha peeps here towards resources that we have which you might be interested in...
And now, to Open Library.
So, what is this thing? A whirlwind tour!
There are a few different ways to describe what it is, but I think the explanation that makes the most sense is “a Wikipedia for Books”.

A “Wikipedia for Books”
A wiki is a website that allows the easy creation and editing of any number of interlinked web pages via a web browser using a simplified markup language or a WYSIWYG text editor.

http://en.wikipedia.org/wiki/Wiki
Open Library is a website that allows the easy creation and editing of any number of interlinked catalog records via a web browser using a simplified markup language or a WYSIWYG text editor.

And a slight tweak specific to Open Library.
One web page for every book ever published.

Our goal is to have one web page for every book ever published. So, at its heart, Open Library is an open, editable bibliographic data repository.
I find it useful to hang discussions of Open Library on to 3 main concepts...
Getting new records...

The project began by collecting bibliographic records from libraries around the world like the Library of Congress, the Internet Archive, Cambridge University and many more. So far, we have about 25 million edition records, millions of authors, millions of book covers, millions of subjects.
This is ImportBot. He gets new catalog records from the Library of Congress and puts them into Open Library every Tuesday. We also import records from Amazon, and from the Internet Archive. ImportBot looks for recently scanned books, and creates new records (or merges them with existing ones) just a few minutes after the record is created on the Internet Archive.
You can see ImportBot working away, just like you can see the Wiki’s edit history for every person who edits something.
Individuals can also add new books with a few details like Title, Author, Publisher and Publish Date. That’s enough for a stub, and then people are invited to add more details.
This is a screenshot of books added – which is available as a filter on the Recent Changes page. You can see we average about 10–20 books per hour, added by individuals around the world.
I am not a librarian. I was on the original team which built flickr.com, that photosharing website. It’s a messy chaotic place from which order emerges organically. I wanted to see if some of the things I learned there could be applied to a library catalog, and riffing of David’s talk on Monday, “Library Data for Fun and Profit, there was no denying that the Open Library catalog is BIG DATA, but like so many other library catalog, its treasures are often obscured.
“Library metadata is diabolically rational.”

Karen Coyle

I’ve enjoyed collaborating with Karen Coyle, advisor to the Open Library, and self-described metadata czar of the project. She’s been profoundly useful helping the team navigate all the data we have, and answering our silly questions.

I was struck by her description of the data we’re dealing with.
I also noticed that loads of library websites start with a search box. The visitor is expected to know what they’re looking for, and there’s no way to browse around.
- There's a presumption of knowledge, not encouragement of exploration
- How do I know what to search for if I don't know what you've got?
- How do I wander the stacks?
I was very pleased to discover the writings of a chap called Alberto Manguel. His thinking on libraries has been deeply influential.
“Books, even after they have been given a shelf and a number, retain a mobility of their own. Left to their own devices, they assemble in unexpected formations; they follow secret rules of similarity, unchronicled genealogies, common interests and themes.”

*Alberto Manguel, The Library at Night*
/Page 163, “The Library as Chance”

So, in the context of diabolical rationality, this excerpt seems appropriate to exposing the depths of the catalog.
“Books, even after they have been given a shelf and a number, retain a mobility of their own. Left to their own devices, they assemble in unexpected formations; they follow secret rules of similarity, unchronicled genealogies, common interests and themes.”

Alberto Manguel, *The Library at Night* 
Page 163, “The Library as Chance”

Here are some other interesting examples...
Catalog as Landscape?

How do you see the shape of a catalog? How can you see its boundaries and edges? How can you move through it differently than search and retrieve?

Photo by me
So, as David talked about on Monday, what treasures lie in wait within this diabolical scheme of description that are waiting to be uncovered? What happens if you blow it up?
Here is a MARC record. Full of data!

Let’s see what happens when you explode Library of Congress Subject Headings. This data isn’t even in Open Library – we borrowed it from loc.gov then pulled out the dynamite...
650 0 $aTeenage boys$vFiction.
650 0 $aBrothers and sisters$vFiction.
650 0 $aPreparatory schools$vFiction.
650 4 $aAlienation in teenagers$vFiction.
650 4 $aTeenage boys$xInterpersonal relations$vFiction.
650 4 $aEmotionally disturbed teenage boys$vFiction.
Teenage boys Fiction
Brothers and sisters Fiction
Preparatory schools Fiction
Alienation in teenagers Fiction
Teenage boys Interpersonal relations Fiction
Emotionally disturbed teenage boys Fiction
Teenage boys, Fiction, Brothers and sisters, Preparatory schools, Alienation in teenagers, Interpersonal relations, Emotionally disturbed teenage boys
Teenage boys, Fiction, Brothers and sisters, Preparatory schools, Alienation in teenagers, Interpersonal relations, Emotionally disturbed teenage boys

Then put a page behind all of them...
Brothers and sisters 2,664 books published between 1640-2011

Publishing History  This graph is interactive. Click and drag to select a range, or click a single year and we'll load the books above.

Looking at the subject page, you can see the Works with the most editions in the top panel, with a handy indicator to tell you if you can read an electronic version....
If I scroll down...we’ve collated all the publish dates of all the editions with that subject.
And, we can also display subjects that are used most often in conjunction with “Brothers and Sisters”, as well as the authors who write most about them, and publishers who publish books about them.
We can also collect subjects together at the author level. Here you can see what sorts of subjects Salinger writes about, each of which links to its page...
Subjects related to J. D. Salinger – note that we’ve retained the Place/Person/Time categories.
Incidentally, my colleague Lance Arthur popped in and updated the Salinger record with a note of his death.
Publishing History This graph is interactive. Click and drag to select a range, or click a single year and we'll load the books above.
Civil War, 1861-1865

Publishing History This graph is interactive. Click and drag to select a range, or click a single year and we'll load the books above.

Related...

SUBJECTS
History, Regimental histories, Biography, Personal narratives, Soldiers, Campaigns, Registers, Personal narratives, Confederate, Religious aspects, Religion, Politics and government, Generals, Genealogy, Juvenile literature, Correspondence, Diaries, Naval operations, Prisoners and prisons, Poetry, Causes

PLACES
United States, Confederate States of America, Virginia, South Carolina, Southern States, North Carolina, Georgia, Texas, Pennsylvania, Tennessee, Missouri, New York (State), Illinois, Louisiana, Ohio, Kentucky, Maryland, Mississippi, Alabama, Massachusetts

PEOPLE

TIMES
Any guesses as to this subject?
La Comète de Halley!
Every 75 or 76 years, people write about it again :)
It’s been great fun exposing this data in new ways, and uncovering all its indiosyncrasies...
Like, tiny variations in data entry.

I mean, is Halley’s Comet really a PLACE?
Wondering about whether or not you could actually stand on the surface of Halley’s Comet... Is that a helpful classification of a place?
Another quick note on data in before I move on...

We’ve been experimenting with a couple of other “surgical” bots, that look across the catalog and connect edition records directly to other services by stamping identifiers from other systems into Open Library. This is a bot written by a developer called Ben Gimpert, that takes a file mapping ISBN to Goodreads IDs, and looks for ISBN matches in OL, then adding the Goodreads ID to those records. This allows us to construct links to Goodreads, and to make the Goodreads ID available through the API.
You can see we’ve added a little widget on the page that connects to Goodreads, if you have an account, you can add our records to your lists on Goodreads. There’s also a LibraryThing ID too, added by a similar batch bot update.

Writing bots to do things like this is the sort of development we’d like to open up to external developers too...
OK. So, we have all this data... Now what?
We have a repository that mostly contains records created by professionals. I find it useful to consider these records as a substrate, something that can be reacted upon.
What if we consider the source Open Library records like that?

Now that we’ve begun to reveal this substrate, how will people react to it? What reactions has it caused so far?
So... Open Library is a virtual space. Its organization isn’t constrained like a physical catalog. In fact, the more connections you can make into one of our “virtual index cards” the more ways people have to discover and navigate its contents.

http://www.flickr.com/photos/brixton/1394845916/
San Francisco Main Library, wallpaper made out of old card catalog files
http://www.flickr.com/photos/ecastro/3089541727/
What if a catalog looks like this? Is crystalline? What if it is unconstrained by the need to sort, say, alphabetically?

From the artist of this image, Jared Tarbell: “Lines like crystals form at perpendicular angles to existing lines. A complex form emerges. 1000 classic computational substrate, color palette stolen from Jackson Pollock: A simple perpendicular growth rule creates intricate city-like structures. The simple rule, the complex results, the enormous potential for modification; this has got to be one of my all time favorite self-discovered algorithms. Lines likes crystals grow on a computational substrate.”
1) Books are for use.
2) Every reader his [or her] book.
3) Every book its reader.
4) Save the time of the User.
5) The library is a growing organism.

So, on the basis of the idea of our current catalog being a substrate, as Ranganathan suggests in his five laws of library science...
1) Books are for use.

2) Every reader his [or her] book.

3) Every book its reader.

4) Save the time of the User.

5) The library is a growing organism.

So, on the basis of the idea of our current catalog being a substrate, as Ranganathan suggests in his five laws of library science...
Handwritten scribbles and scrawls; annotations; corrections
What happens when you introduce turbulence into the catalog? Here are a few examples of the sorts of edits we’re seeing... at a rate of about 100,000 edits per month.

http://www.flickr.com/photos/rreis/4859722551/sizes/l/
What happens when you introduce turbulence into the catalog? Here are a few examples of the sorts of edits we’re seeing... at a rate of about 100,000 edits per month.

if you don’t stimulate an organism, it atrophies

http://www.flickr.com/photos/rreis/4859722551/sizes/l/
Edits

Now, I’m going to fly through the sorts of reactions I’ve noticed...
It’s been a thrill to see new non-English records coming in...
本気で書きたい人の 小説「超」入門
Published 2005 1 29 by たなか みよし in 埼玉県さいたま市浦和区.

About the Book
本気で小説を書きたい人のために。
さらに力を発揮しようとしている人のために。

Table of Contents
1. 心得　書きますか、書きませんか
2. 探求　伝えたこと、言いたいこと
3. 前進　伝わるかどうかはあなた次第
4. 克服　こんな書き方では読んでもらえない！
5. 精進　自分の中の日本語を生かして
6. 継続　書く力をさらに高めるために
7. 習得　題材と文章は永遠のパートナー
8. 実践　覚悟はいいですか！
9. 激励　もっと書きたい人よ
補記　さまざまな表現を味わおう
書き続けているみなさんの覚悟と方法
あとがきに代えて——最後まで書け、声に出て読み直せ
ご案内——赤ペン編集者があなたの原稿に未入れします！

The Physical Object
ΠΕΣΟ ΒΟΥΒΟΥ ΚΟΚΑΛΑΚΟΤΩΡΙ

By Γιάννης Φαρσάρης

About the Book
Η συλλογή μικρών ιστοριών "Εβδόμη εσπερινή" του Γιάννη Φαρσάρη διανέμεται ελεύθερα στο Διαδίκτυο σε μορφή ψηφιακού βιβλίου στη σελίδα http://www.openbook.gr
/2010/10/ebdomh-esperinh-mikres-istories.html

SUBJECTS
Αισθήματα, ebooks, free ebooks

There is only 1 edition record, so we'll show it here. Add another?

Εβδόμη εσπερινή
Μικρές ιστορίες
Published 2010 by Openbook in Ελλάδα.
Written in Greek.

About the Book

Table of Contents
1 Μήνας μπαίνει, μήνας βγαίνει - 07 -
2 Θα κάτσω στίτι, θε' αράξω στίτι - 10 -
Here’s an example of what I would call a subject domain expert. You can see from his recent edits that he’s interested in zoology...
Domain Knowledge:

“A good introduction to the zoo plankton. Illustrated throughout.”

Clicking through to his edits, he’s noting which books are good for what, as well as adding additional metadata descriptors.
There’s a chap called David Murray who has used Open Library to catalog his father’s work developing and publishing marriage records for the USA
New Subjects:
Created by a relative of the genealogist who wrote all these publications.

Louisiana surname index
Murray, David A., Nicholas Russell Murray...
If you look at the edit history of this record, there’s been an update made by someone connected to the author.
Hi, :) my grandparents live in Miss Nelson's former home. I added her year of death, birthdate, and everything I know about her.

You’ll notice that the comment left on the latest edit is by someone connected to the author.
Authors appear...
Who better to maintain a catalog record than the author herself?
Arizona's territorial newspapers

By Sam S. Webb

About the Book
This book describes the early territorial newspapers in Arizona. It has chapters on the major newspapers of that time period and provides lists of the owners and editors along with political affiliations. It contains historical photographs of the newspapers and reprints of stories.

SUBJECTS
Journalism, History, Frontier and pioneer life

PLACES
Arizona, Tombstone, Phoenix, PrescottTucson

TIMES
20th century, 19th century

There is only 1 edition record, so we'll show it here. Add another?

Arizona's territorial newspapers
a brief history of frontier journalism in Arizona
by S.S. Webb.
Published 1997 by S.S. Webb in Borrego Springs, Calif.
Written in English.

Edition Notes
Includes bibliographical references (p. 46).

You can see a new description on this work, and...
“I am the author of this book and provided a brief synopsis of what is contained in this publication.”
Deep space 1 edition
By Stephen Whitfield

About the Book
Space junkies and armchair astronauts are provided with comprehensive, handy references for a variety of space-related missions, vehicles, and concepts in this pocket-sized series. Compiled with the cooperation of NASA, each topic-specific orbiters, probes, and experimental equipment are featured in each equipment-specific reference.

Flight data and images for all of the unmanned deep space probes, including Cassini, Deep Space One, Galileo, Huygens, NEAR, Pioneer 10 and 11, Stardust, and Voyager 1 and 2, are featured in this examination of astronomers' attempts to understand the outer reaches of the universe.

There is only 1 edition record, so we'll show it here. Add another?

Deep space
by Stephen Whitfield.
Published 2005 by Apogee Books in Burlington, Ont.
Written in English.

Wednesday, October 27, 2010
Deep space

By Stephen Whitfield

Space junkies and armchair astronauts are provided with comprehensive, handy references for a variety of space-related missions, vehicles, and concepts in this pocket-sized series. Compiled with the cooperation of NASA, each topic-specific reference features relevant statistics, photographs, and the stories behind each project. Books on manned missions include crew photographs, information on patches and equipment, and flight statistics such as time in space, distance traveled, and mission objectives. Photographs and statistics for launch vehicles, orbiters, probes, and experimental equipment are featured in each equipment-specific reference.

Flight data and images for all of the unmanned deep space probes, including Cassini, Deep Space One, Galileo, Huygens, NEAR, Pioneer 10 and 11, Stardust, and Voyager 1 and 2, are featured in this examination of astronomers' attempts to understand the outer reaches of the universe.

There is only 1 edition record, so we'll show it here. Add another?

Deep space
by Stephen Whitfield

Published 2005 by Apogee Books in Burlington, Ont.
Written in English.
Duplicate authors (and editions) are an issue... This is an example search for author records with “author” in their names... you can see the variety of ways that catalogers have noted unknown authors...

http://openlibrary.org/search?author=author
Author merge is an example of one of the tools we’ve built to help people clean up the data. There are a ton of cases where tiny differences between data entry practices, and that’s something hard for a computer to spot. But humans are very good at it. This author, Danny Howell, has searched for himself, found duplicate entries, and merged them himself.
He’s taking care of his own records...
And has added several new records and covers for his own books.
History

History is an essential part of any wiki. You can see any/all changes to every record, and even revert changes if you make a mistake, or want to clean something up quickly.

A big part of the redesigned site which launched in May this year was to surface the general activity of the site more prominently, in the hope that it would indicate that there was a lot going on.
Detailed history list, from which you can select two different versions to compare. If I choose the first version, and the most recent, I can see all the edits that have been made.

<table>
<thead>
<tr>
<th>REVISION</th>
<th>WHEN</th>
<th>WHO</th>
<th>¿QUÉ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>15 seconds ago</td>
<td>George</td>
<td>Added second author; title case; ToC tidy</td>
</tr>
<tr>
<td>8</td>
<td>2 hours ago</td>
<td>ImportBot</td>
<td>Found a matching record from Library of Congress.</td>
</tr>
<tr>
<td>7</td>
<td>1 day ago</td>
<td>ToCBot</td>
<td>Adding Table of Contents</td>
</tr>
<tr>
<td>6</td>
<td>August 4, 2010</td>
<td>IdentifierBot</td>
<td>added LibraryThing ID</td>
</tr>
<tr>
<td>5</td>
<td>May 5, 2010</td>
<td>ImportBot</td>
<td>Added new cover</td>
</tr>
<tr>
<td>3</td>
<td>May 5, 2010</td>
<td>ImportBot</td>
<td>add scanned books from the Internet Archive</td>
</tr>
<tr>
<td>3</td>
<td>May 5, 2010</td>
<td>ImportBot</td>
<td>add scanned books from the Internet Archive</td>
</tr>
<tr>
<td>2</td>
<td>December 14, 2009</td>
<td>WorkBot</td>
<td>link works</td>
</tr>
<tr>
<td>1</td>
<td>April 1, 2008</td>
<td></td>
<td>Initial record created, from Scriblio MARC record.</td>
</tr>
</tbody>
</table>
The full diff view
In addition to browsing the catalog, we have a SOLR search engine.
you can do a simple keyword search, or focus on specific fields like title, author or subject.
You can use search facets to zero in fairly quickly on books you can read online now, in this example, I found 4 ebooks about Wellington.
Of course, if I’m not sure, I can always broaden my search again and just browse...
We’ve just released a rebuild of full text search too, and there are brand new Search Inside APIs on the site, hot off the press.
New Zealand Cities by Charles Wilson
Published in 1900 by Whitcombe & Tombs • 68 pages • snippets in situ

... . . . . . Frontispiece
Oriental Bay . . . . . . . . . . . . 11
Willis Street . . . . . . . . . . 13
Bank Corner . . . . . . . . . . . 16
Thorndon...

..., and leaving Oriental Bay on the left, makes straight for her berth at the Queen's Wharf. At night the circling town lights provide a scene of fairy-like beauty...

... of the average pedestrian, is from the Oriental Bay tram terminus, "round the rocks" as the local phrase goes, to the Wellington Patent Slip and Kilburnie. All...

Bulletin – United States National Museum. by unknown author
Published in 1877 by Smithsonian Institution Press, [etc.], for sale by the Supt. of Docs., U.S. Govt Print. Off. • 908 pages • snippets in situ

..., vol. 16, p. 296, fig. 1, 1935 (Oriental Bay). HeptraneMas macdonaldi (Ogilby) Whitley, Rec. Australian
This is a view of the search snippets in place on the pages, so you can do a quick scan for context before you decide to open the bookreader.
I can open the book and be reading it within seconds.

(Again, this is the new bookreader, coming out soon!)
I sometimes have to pinch myself to remember that even though we’re a wiki, and open, and have all these nice interfaces to explore the catalog, what we’re REALLY trying to do is connect people to books. To help people find something to read. So, we connect to actual books wherever we can.
Accessible book 657,305 works / 1,078,835 ebooks published between 1099 & 10000.

Publishing History  This graph charts editions published on this subject. Click to view a single year, or drag across a range.

Over 1 million free books! Available now!
There’s also a large collection of works available to the print-disabled community through the NLS program at the Library of Congress. These books are available in DAISY format to anyone registered for a key from the NLS, which will open these DAISYs on a text to speech device.
We’ve also connected to Overdrive, which is a system that manages loans for various libraries around the world. All of these Overdrive titles have links that send you off to Overdrive to borrow an eBook.
Snapshot of the various combinations of links we can provide to get you to books... We’ve worked with Overdrive in the states to get eBook versions of things for loans... We’re hoping to make the vendors you can buy from more dynamic, and open up the sources for online free texts. Right now, it’s just the Internet Archive texts that we link to in full.

That’s a glimpse of the sorts of Activity we’ve seen so far... now on to Data Out.
As I mentioned, I spoke to Chris a long time ago about Open Library and Koha beginning to work together.

I wanted to give a quick overview of the ways you can get data back out of Open Library, something we’re looking to improve on constantly, and focus on into next year.
I’m sorry to say that if we’re completely honest, we’re open source by name only... It’s been really inspiring to witness KohaCon though. I’m incredibly impressed by the strength and productivity of the Koha development community. Something to aspire to, to be sure!
We certainly have our code online at github, but we rarely receive patches. I’m OK with this, at least for now.
We’ve had some awesome patches for the bookreader though – like the Thumbnail view, submitted by Stephanie at the California Digital Library.
The Library of Congress is using our Bookreader on read.gov. There are quite a few other examples of the IA Bookreader out there on the web. Hopefully the redesign (with touch interactions etc) will attract new people too...
I think it’s much more practical for us to consider ourselves as a free, open, editable bibliographic repository.
Bulk Download

Open Library provides dump of entire edition and author records in JSON format. If you want to download everything, please use this instead of our API.

- All Records
- Authors
- Editions
- Works

Today, you can download the whole catalog. But, that’s a HUGE dataset, and hard to work with.
Or, you can download individual editions, works or author records as RDF, JSON, or create a Wikipedia citation that you can just paste straight in.
### History

<table>
<thead>
<tr>
<th>Date</th>
<th>Action Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 29, 2010</td>
<td>Edited by WorkBot · merge works</td>
</tr>
<tr>
<td>September 27, 2010</td>
<td>Edited by Librarian M · merge authors</td>
</tr>
<tr>
<td>July 13, 2010</td>
<td>Edited by EdwardBot · tidy contributions list, add notes, improve work titles</td>
</tr>
<tr>
<td>April 13, 2010</td>
<td>Edited by Open Library Bot · Linked existing covers to the edition.</td>
</tr>
<tr>
<td>April 11, 2010</td>
<td>Created by ImportBot · Initial record created, from Internet Archive MARC record.</td>
</tr>
<tr>
<td>Date</td>
<td>Action Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>September 29, 2010</td>
<td>Edited by WorkBot, merge works</td>
</tr>
<tr>
<td>September 27, 2010</td>
<td>Edited by Librarian M, merge authors</td>
</tr>
<tr>
<td>July 13, 2010</td>
<td>Edited by EdwardBot, tidy contributions list, add notes, improve work titles</td>
</tr>
<tr>
<td>April 13, 2010</td>
<td>Edited by Open Library Bot, Linked existing covers to the edition.</td>
</tr>
<tr>
<td>April 11, 2010</td>
<td>Created by ImportBot, Initial record created, from Internet Archive MARC record.</td>
</tr>
</tbody>
</table>
LEADER: 00670cam a22002291 4500
001 61011228
003 DLC
005 20100823182117.0
008 750716s1961 nyua 000 0 eng
010  $a 61011228
035  $a(OLC)1451322
040  $aDLC$CFOS$dOCOLC$dPJo$dDLC
042  $apremarc
050 00 $aTX840.B5$bS55
082  $a641.5
100 1 $aSeranne, Ann, $d1914-
245 14 $aThe blender cookbook, $cby Ann Seranne [pseud.] and Eileen Gaden. Photographic illus. by the authors.
250  $a[1st ed.]
300  $a288 p. $bilus. $c22 cm.
650 0 $aCooking.
700 1 $aGaden, Eileen, $ejoint author.
There’s a copy of everything we’re using on the Internet Archive too.
We've been working to improve our API offering, and have released several new public APIs in the past few months.

In addition to simply getting book records back, you can now also ping for subject-related data and try out our brand new Full Text Search API.

We're also working on a new Lists feature, and we actually wrote the APIs first, and then built the feature off that.
One of quite a few examples of Open Library in the wild includes the National Library of Australia’s new search engine, Trove.
You can see there that there are links to Open Library books wherever one can be sourced.

There are a growing number of sites making use of Open Library data... and that’s what we’re all about – data in, data out. The more interconnections we can make with other systems, the easier it will be for people to land where they want to go inside Open Library.
To conclude...

Some Koha ideas

If I were able to come to the hackfest – and I’m really sorry I won’t be there – I would be looking to try a few things...
I’m so pleased to be here to begin the conversation of future integrations between Koha and Open Library. I wonder if these few ideas might be good places to start...

(discuss ideas)

And with that...