the amazing site that is

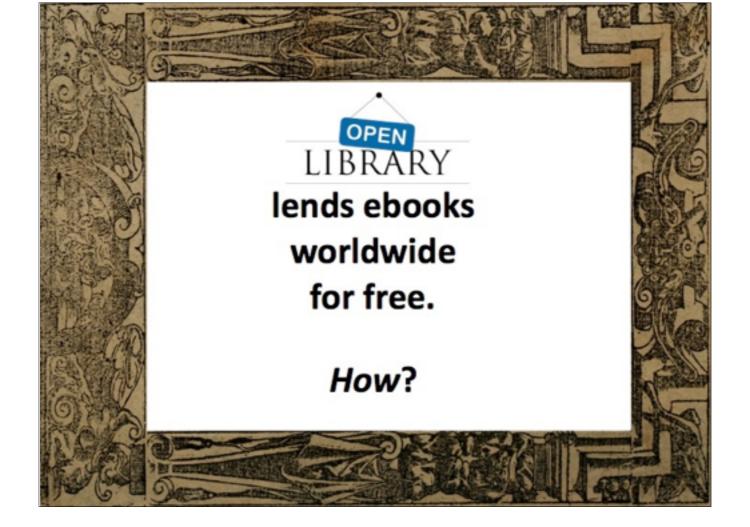


This is a talk I did to Internet Archive Staff about the Open Library project.

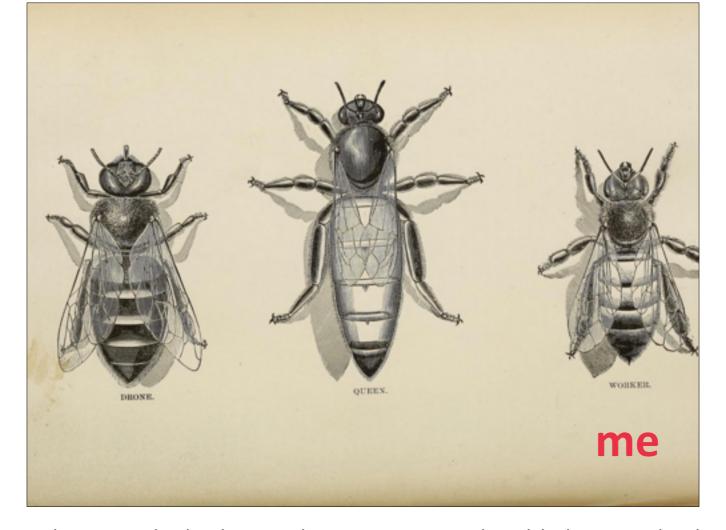


quick backgrounder

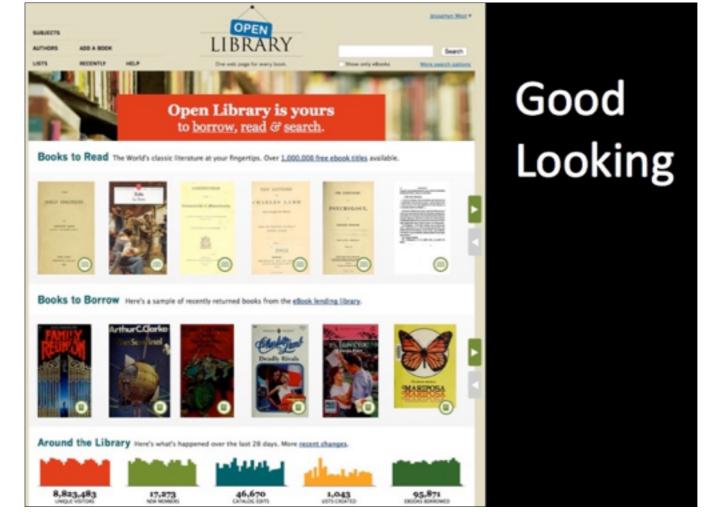
This is just a bulleted point list of things that Open Library is known for. I've been working with them on and off since 2007 but as an employee since 2014.



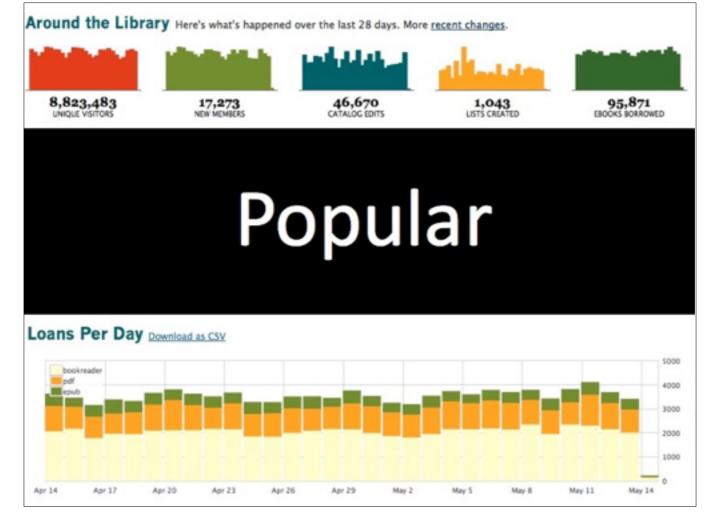
This is our tagline.



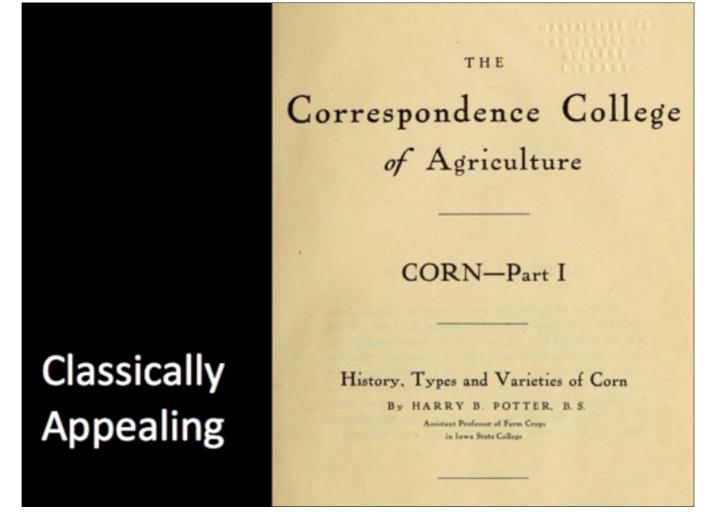
I am not the project manager, I am mostly just a worker bee but I am the one person at IA who solely does OL and understands the entire project at a decent level.



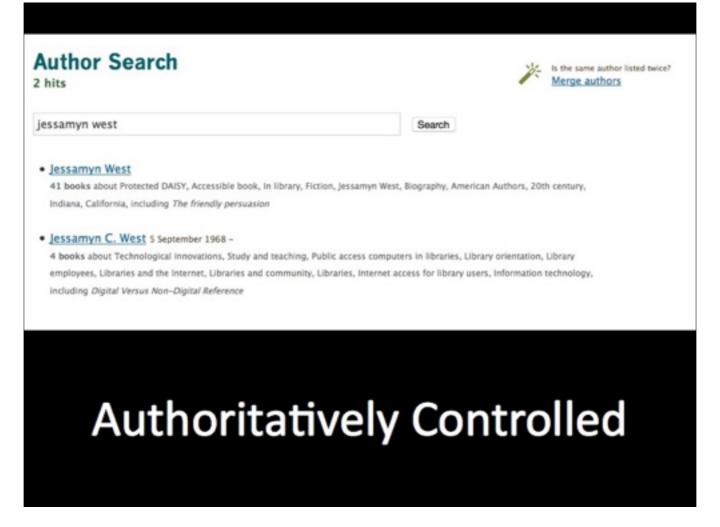
Open Library is a nice looking site. The Internet Archive site is too, since the makeover, but this wasn't always true.



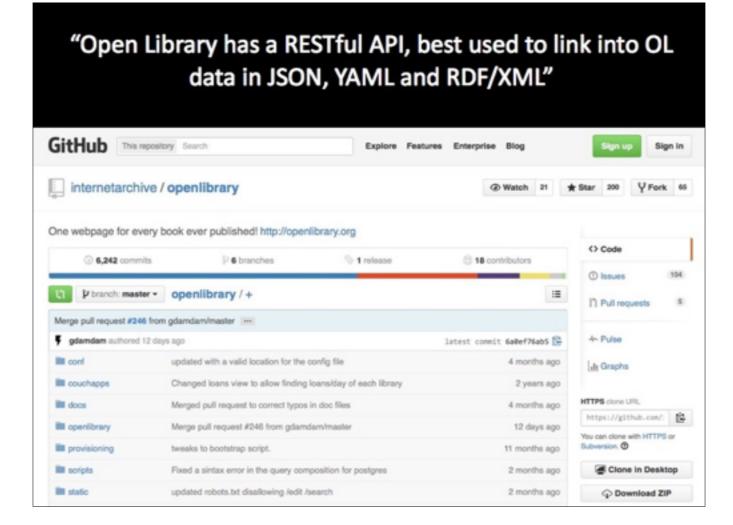
We get a lot of traffic. 300,000 unique IPs per day 3-4000 loans per day.



This is sort of a joke slide since when you search Harry Potter you will find the JK Rowling books but also a lot of stuff like this. We have over a million free ebooks available to lend.



We use authority control for authors which means each Author has a unique ID in our database and we don't have like seventy entries for various spellings of their names (ideally, some of our data is not the cleanest, but we fix it when we see it)



We also have the code available on GitHub and an API people can mess around with. I don't know the nitty gritty about this stuff but I loosely understand how it works.

Add a book to Open Library We require a minimum set of fields to create a new record. Title	These are those fields.	Users can	
Author Like, "Agatha Christie" or "Jean-Paul Sartre," We'll also do a live	theck to see if we already know the author.	add records.	
Who is the publisher?			
When was it published? The year it was published is plenty.			
And optionally, an ID number — like an ISBN — v	Editing		Delate Record
Add Cancel	Title use title falente to act a sateria. "Inquired field		
	Justin Morgan Had a Horse		
	Author Marguerite Henry	Add another author?	
	What's It About? ADD-DICERTS U	nor5	
	How would you describe this book	There's no wrong prower here.	
Heave see	B / 2400 HES		?
Users can			
edit records.	Preview		
	formation and and		
		commun. For example: cheese fumar Empley acachology organ (Horse), Morgan horse, Folklore, Legends and stories, In library, J	luvenile litera

It's also basically a wiki so users can add AND edit our records. Not all the information can be changed by users, but most of it can.

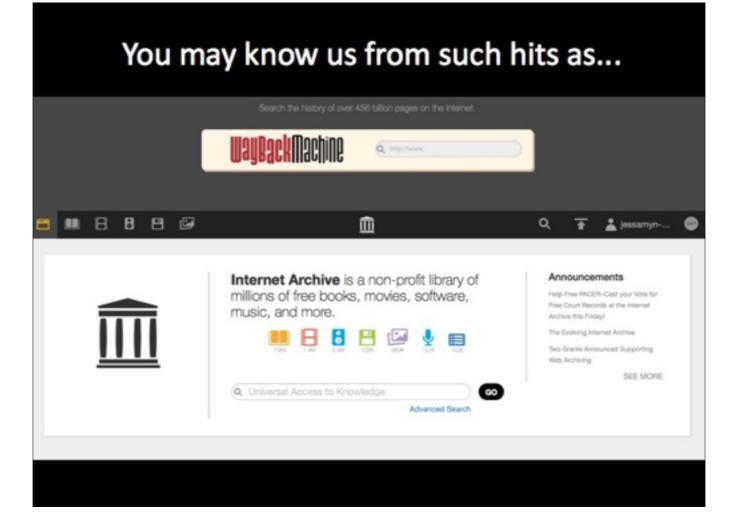


So, let's go backwards in the story a little...

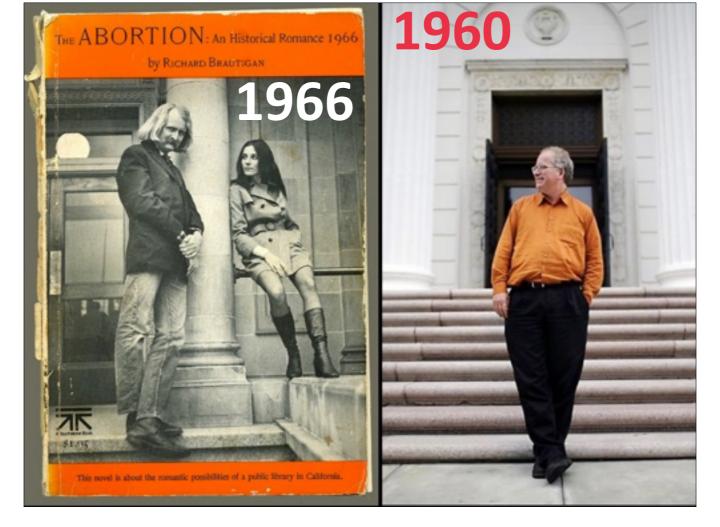


some history

and talk about where it came from, sort of



If you're not familiar with the Internet Archive you may know our main popular tool the Wayback Machine. Archived websites going back to the birth of the web. A very cool tool.



The Abortion is a book by Richard Brautigan (only sort of about an Abortion) that was written in 1966. It's fiction and had the idea of a library in the heart of San Francisco that was open 24 hours and anyone who wrote a book could bring it there and put it on the shelf and make it available to others. It's the book that made me want to be a librarian, at THAT library. This is Brewster, he was born in 1960. The Internet Archive, also in San Francisco is the closest thing we have to a real version of the Brautigan Library.



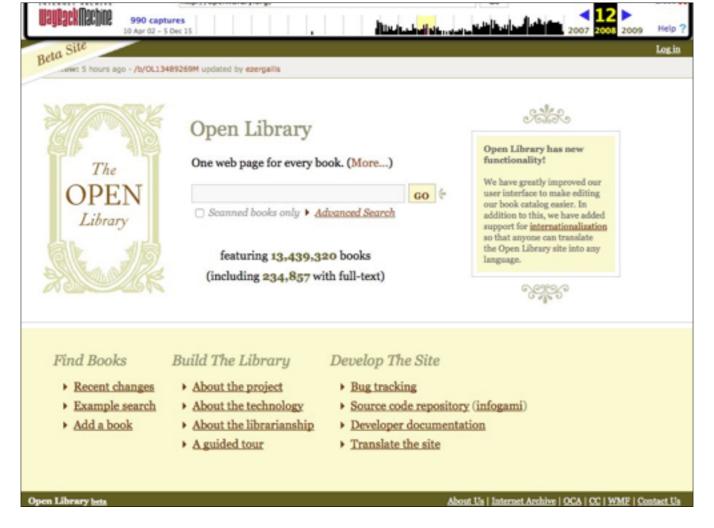
Back in Febrary 2007 there was a meeting at the Archive to talk about this idea. You may recognize some faces, notably Aaron Swartz who was really jazzed about this idea.

RAW THOUGHT by Aaron Swartz Announcing the Open Library Early this year, when I left my job at Wired Digital, I thought I could look forward to months of lounging around San Francisco, reading books on the beach and drinking fine champagne and eating foir gras. Then I got a phone call. Brewster Kahle of the Internet Archive was thinking of pursuing a project that I'd been trying to do literally for years. I thought long and hard about it and realized I couldn't pass this opportunity up. So I put aside my dreams of lavish living and once again threw myself into my work. Just as well, I suppose, since San Francisco's beaches are freezing cold, champagne has a disgusting taste, and fore gras is even worse. I thought of the smartest programmers and designers I knew and gave them a ring, sat down for coffee with them, threatened to fly out to their homes and knock on their doors. In the end, we got together an amazing group of people -- all sworn to secrecy of course - and in the past few months we've put together what's probably the biggest project I ever worked on. So today I'm extraordinarily proud to announce the Open Library project. Our goal is to build the world's greatest library, then put it up on the Internet free for all to use and edit. Books are the place you go when you have something you want to share with the world - our planet's cultural legacy. And never has there been a higger attempt to bring them all together. I hope you'll take a look and let me know what you think. And if this project excites you the way it excites me, I hope you'll join us. · Visit the Open Library site You should follow me on twitter here. July 16, 2007

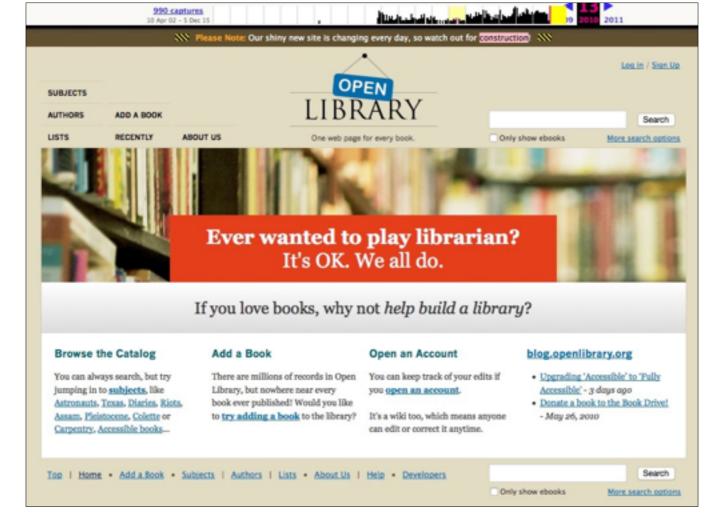
He wrote about it on his blog and this meeting was a get-to-know-you meeting to talk about making it a reality.



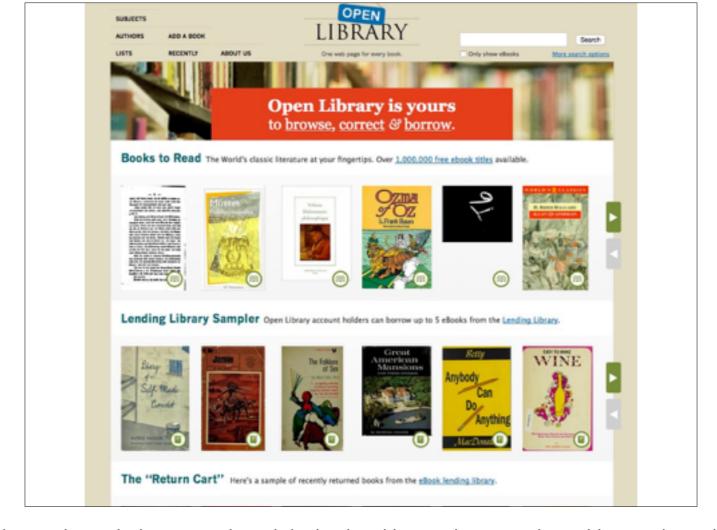
With some funding and some hard work there was a beta version of it that launched in 2005. Mostly just pointing to classic books and letting people browse them with the BookReader which was a completely amazing way to interact with books online at the time.



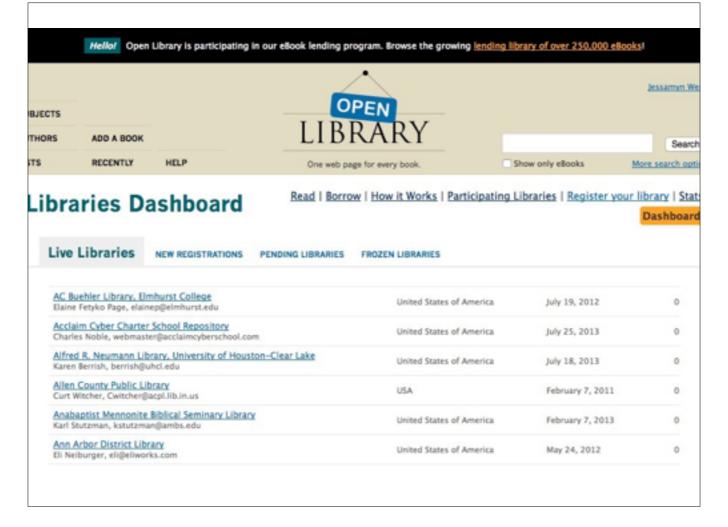
With some more funding and some coding, the site transformed into something where people could interact with a lot of books that were organized in a structured way. This is 2008.



By 2010 the site was more or less in its final form.



... and looked more or less like it does today with the carousels and the lending library where people could not only read books but BORROW them.



At the time, the plan was that brick and mortar libraries would join. They'd share content with us and we'd share content with them. There is some cruft. The libraries program specifically had libraries register and then OL was available specifically to their IP range. They'd send us a book as "registration"

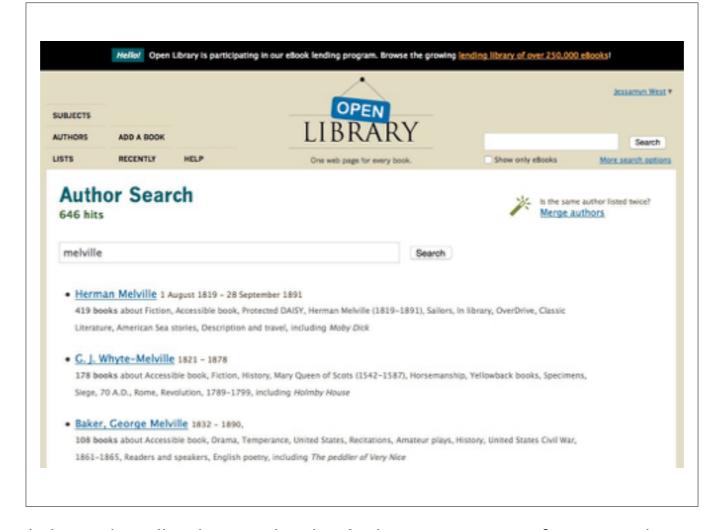


In 2011 this was big news. Brewster and the IA team had made inroads at every state library in the country and this gave IA the freedom to open up lending to all the US states, removing the IP restrictions. Later the books became lendable to anyone in the world.

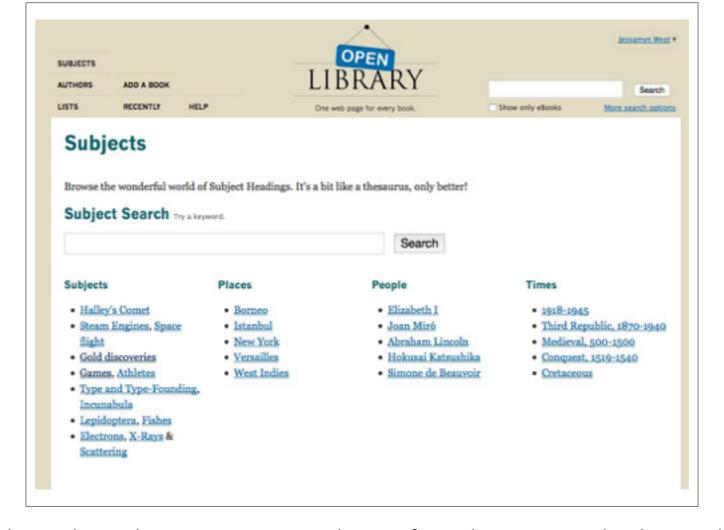


librarianship

And it's not just a big file cabinet of books, there is librarianship going on there.



I talked earlier about authority control, this works well and means that there's also one page on OL for every author.



We have subject headings though they are looser than just DDC or LCSH. There are facets that you can explore by, time, language, locations.



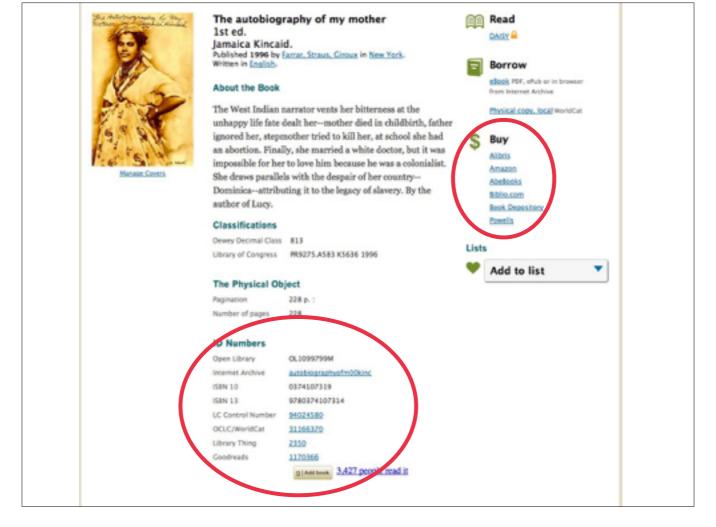
The Marc records are accessible individually or via bulk downloading from the Archive.



We also have pages for publishers which is good news bad news since especially for older publishers, our page on OL shows up higher than other listings for the publisher, causing a lot of weird emails.



While there is a page for every book, we also link editions of a book to the book's master page.

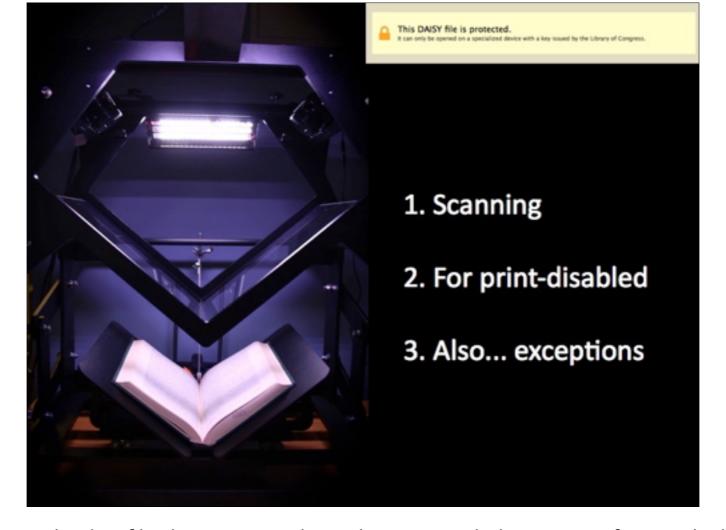


If you look at a page for a specific edition, there are links where you can go buy the thing (if you are impatient) or find it other places. The link to the Internet Archive record is the important thing that makes a book readable or lendable. The digital content itself is at the Archive.



basic mechanics

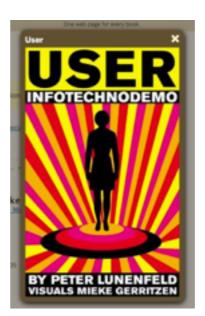
So here's how a loan works, in a general loose sense.



We get records from a lot of places. We do a lot of book scanning. We have a legit reason which is scanning for print disabled which is a legit reason to "format shift" titles.

Users can...

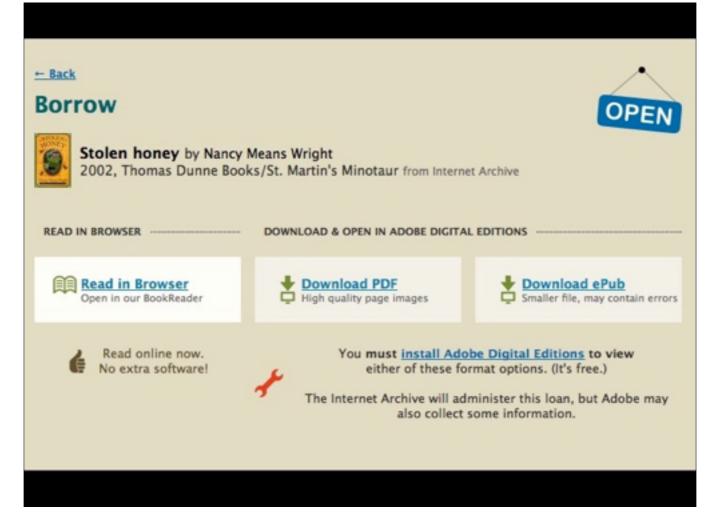
- Read and Borrow
- Waitlist items
- Add or Edit data incl. covers
- Make lists
- Interact with API



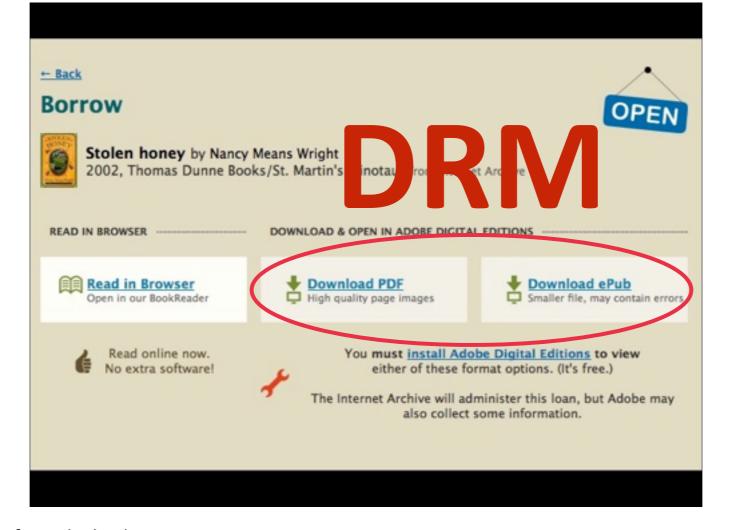
User can do a bunch of things with these records.



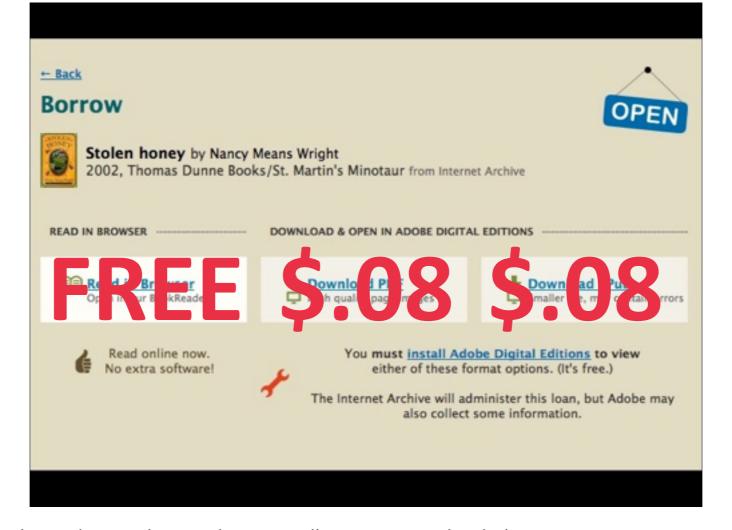
People using Open Library have lending options. DAISY users need a special key which they get from NLB (which is complicated, NLB disavows knowledge of that, we'd like to fix that) which open these books on a BARD reader. We do not have audio books, just DAISY ones.



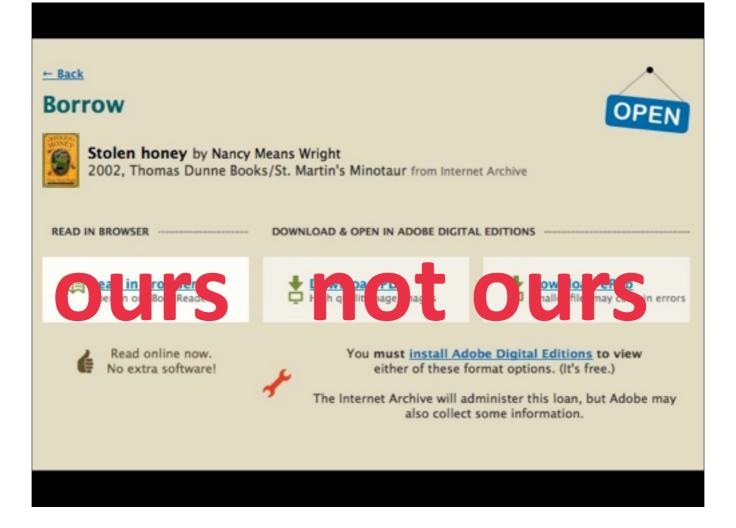
When you say you'd like to borrow a book, you go to this page. The PDFs are direct scans. The EPUBs are the automatically generated files. They are okay but when they say "may contain errors" they mean "definitely contain errors" for items with a lot of images or formatting.



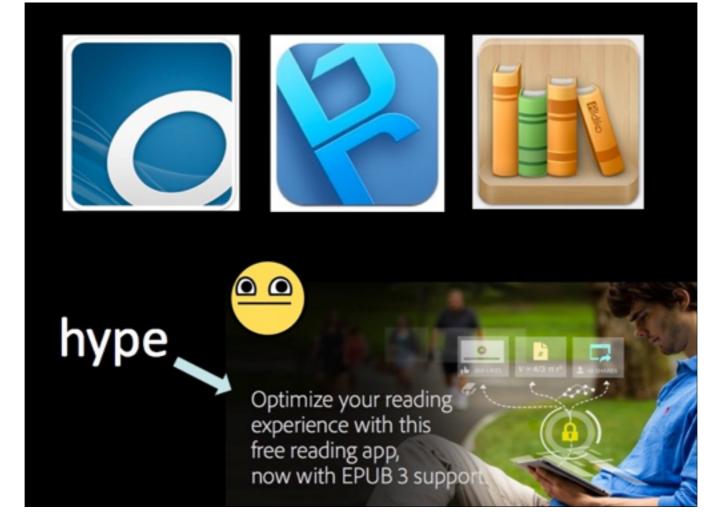
The PDFs and EPUBs use DRM to enforce the lending criteria



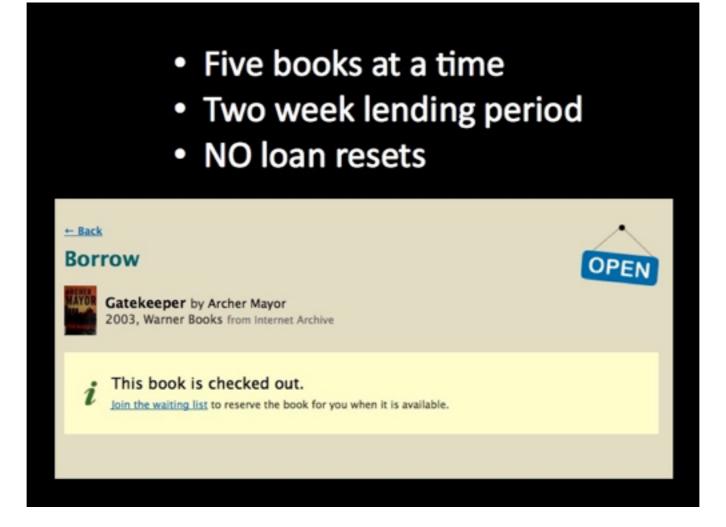
Reading in our browser costs us nothing. The DRM licenses have a small cost associated with them.



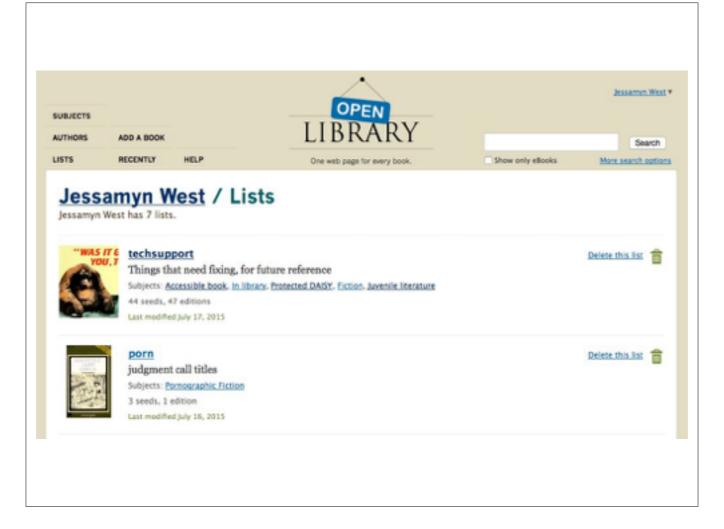
The DRMed licenses are regulated through an Adobe License server which I don't know much about.



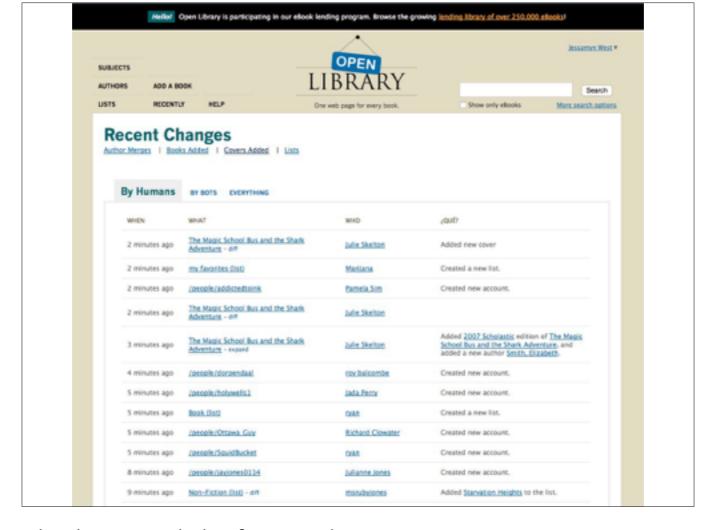
The ebook reader environment is a mess. Many ebook readers interact with Adobe's DRM but they don't all have the same featuresets even though they promise users the moon. It's not optimal for our users but it does give them options.



The rules are really simple. And they are enforced via software.



Users can also make lists to keep track of things like books they've read since we don't keep that information about users.



We can track what people are doing within the system which is fun to watch sometimes.

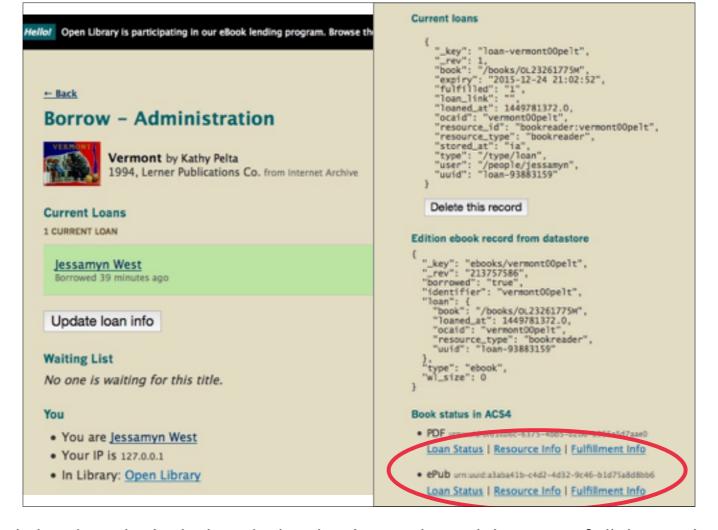


We don't really talk about "The C Word" and I won't here either.



the back end

From an admin perspective, the tools are pretty good.



We can see who has borrowed a book, how long they've had it, whether there's a waitlist and the status of all the past loans for a subject.

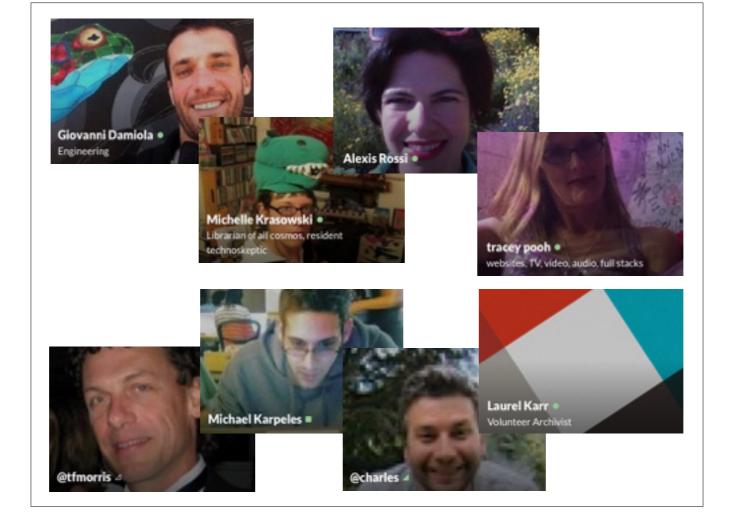
H	Open Lil	vary is participa	eting in o	ur ellook lending progra	n. Browse t	he growing lends	ng library of over 250,000	tellookst
SUBJECTS AUTHORS AS	0 A BOOK			LIBR		,		Joseph Seech
LISTS RE	CENTU	HELP		One web page for	every book.		Show only ellooks	More assess, spriors
People				em. Words Solt Imports. •	Scapha I II	naext.atore truse	et memosche	
Status: acti	WE SHARE	account in	and passw	ont read areal			16	ck and even all edits
Registered on Feb. Activated on Feb			1.87					login as this user
Name:	Jessamyn W	lest - view profi	tle					
Email Address:	jessamyn)	gmail.com		Change				
Reset Password				Ohenge				
Box	No Make 8	M.						
# Edits	5364							
Member Since:	February 23	, 2011						
Last Login:	41 minutes	ago						
Tags:	alumnus	beta-tester	staff	trusted-user				
Loans								
1 CURRENT LOAN				LOAN EXPIRES		WHAT	WHO	ACTIONS
D.	Vermont Borrowed 43	minutes ago		December 24, 20	15	BookReader	(people/jessamys	debug_info
Waiting Lo								
Edit Histo	ry 5364.60	ta.						

We can look at a user's page and see what they have on hold, how long they've been with us and what edits they've made.

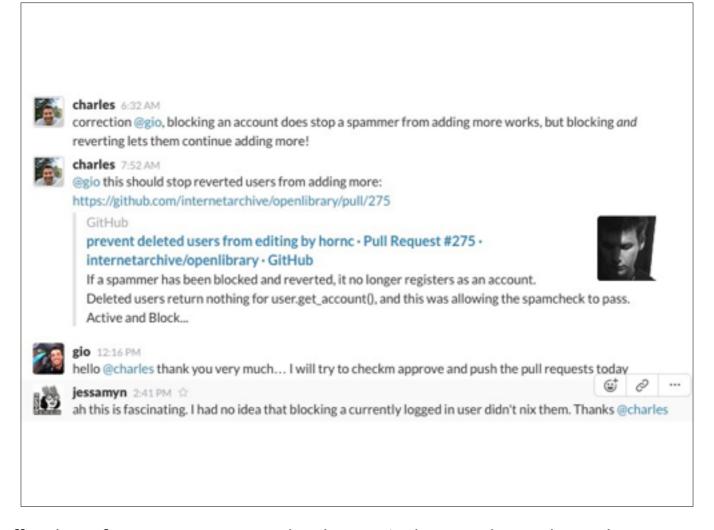


keeping Open Library running

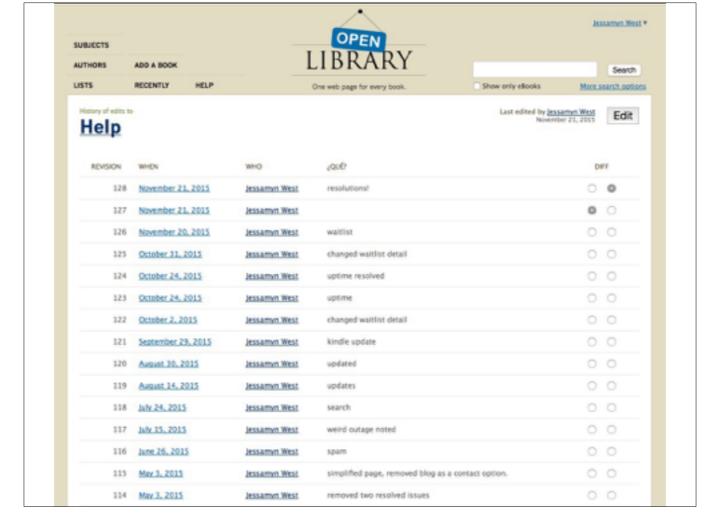
There are a lot of parts that go into keeping OpenLibrary up and running.



I work with a team of staff (top row) and volunteers (bottom row) who all do small parts of keeping thigns working. I'm in the lead email support person, mainly, but I also do a lot of volunteer coordination.



We have a slack channel (one for staff and one for communication with volunteers) where we do our day to day communication



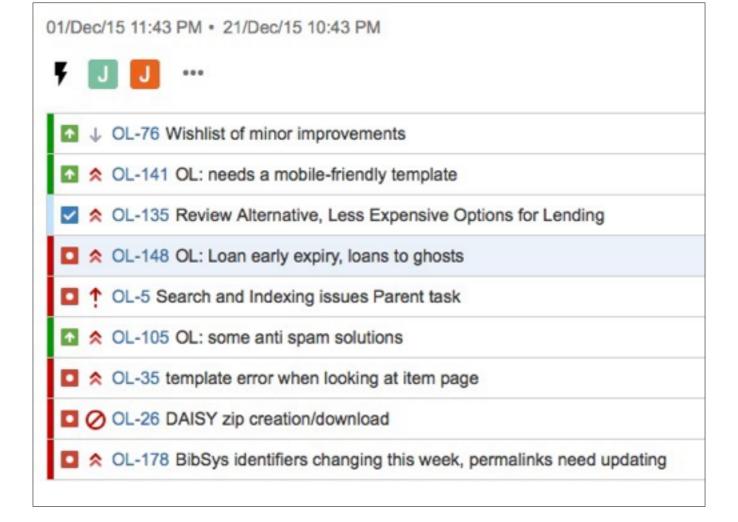
Our help pages are also wiki-editable by staff so we can alert users when there is downtime or an outage or when we've found a bug. We keep them pretty current.

	Hello! Open	Library is particip	sating in our effock lending proj	yam. Browse the growing b	ending library of over 210,000	etioolal
SUBJECTS			OF	EN		Joseph Med *
AUTHORS	ADD A BOOK		LIBE	KARY		Search
LISTS	RECENTLY	HELP	One web page	for every book.	Show only ellooks	More search potions
	t's the p		1? Questions (FAQ) to see	if your question is a	nswered there. Thank y	ou.
Your Er	nail Address					
jessam	yn@gmail.com	1		(We'll need this	s if you want a reply)	
Topic						
Select			0			
Your qu	estion					
If you e	ncounter an er		please cut and paste it provide the url of the o			e book.
Yes I Hen	e • Adda.Book	• Subjects I &	uthers Lists • About lis	Hela • Developers	Show only ellocits	More search settlers
Arou The Libra	Among S.E. cools		Jane Sandberg updated Sustainable Assiculture 1 minute ago	Laura Vinance updated Katasa 1 hour ago	lane Sandberg updated Asticulture, and developmen 2 minutes ago	* (Po

And users email us, a lot. Every page on OL has a little flag in the bottom that encourages users to contact us if there is an issue.

Open Library				
Support Data	Jan '16	2015	Dec '15	Nov '15
Outgoing	558	6900	464	454
Incoming	656	8101	484	564
Books Lent	113,650	1,075,149	97,259	95,705
Incoming by category				
Borrowing books	188	2012	153	109
Buying Books				
Contact an Author				
Deletion request	55	819	27	79
Dev/API	14	174	9	10
Editing Problem	41	329	23	16
Login trouble	34	599	19	32
Other	282	3468	218	266
Spam report	1	154	4	6
Title Mismatch	31	301	19	32
Waitlist Issues	10	245	12	14

We answer between 500 and 700 emails most months and circulate about 100,000 items. You can see how our email load breaks down.



We also keep track of bugs in Jira.

Ol-tech mailing list administration **General Options Section** Configuration Categories Other Administrative Activities · Tend to pending moderator requests [General Options] Privacy options... Go to the general list information page Passwords Bounce processing Edit the public HTML pages and text files Language options Archiving Options · Go to list archives Membership Management... Mail<->News gateways Non-digest options Auto-responder · Delete this mailing list (requires confirmation) Content filtering Digest options Topics Logout Make your changes in the following section, then submit them using the Submit Your Changes button below. **General Options** Fundamental list characteristics, including descriptive info and basic behaviors.

And there are a few mailing lists which are low traffic but have user-helping-user levels of discussion.

tech challenges

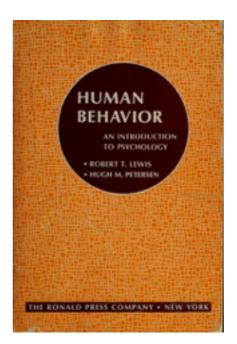
- Search keeping index fresh, loan status correct, fixing bugs
- Data clean & updated metadata, records ingestion, user contributed content
- Alerts response system for updates and outages
- Process for repair and updates - getting from Jira to done



We do have some challenges and things we wish were working better. These are the tech challenges.

human challenges

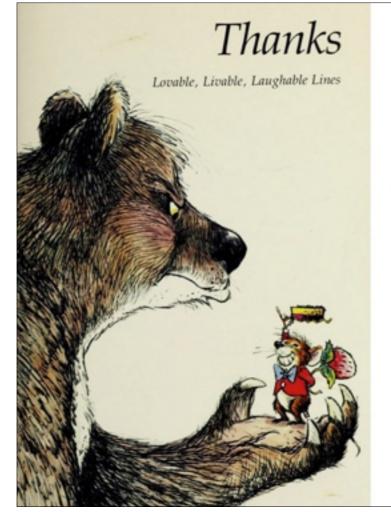
- Setting expectations of users, devs, staff, volunteers
- Harmony with IA lending
- Building a community of users and contributors
- Managing a shared inbox



And some of the challenges are just working with other people and getting things smoothly functioning.



The more we get out and about the more great tools can be built with the mass of content we've made available. Here is my favorite thing, a Flickr photoset of 5,000,000 images that came form public domain books. Free for the using.



Jessamyn West

jessamyn@gmail.com

@jessamyn

Thanks for reading this far. Feel free to contact me if you have questions.